Noisy Compressive Sampling Based on Block-Sparse Tensors: Performance Limits and Beamforming Techniques

Rémy Boyer and Martin Haardt, *Senior Member, IEEE*

***Abstract*—Compressive sampling (CS) is an emerging research area for the acquisition of sparse signals at a rate lower than the Shannon sampling rate. Recently, CS has been extended to the challenging problem of multidimensional data acquisition. In this context, block-sparse core tensors have been introduced as the natural multidimensional extension of block-sparse vectors. The $(M_1, \ldots, M_Q)$ block sparsity for a tensor assumes that Q support sets, characterized by M_q indices corresponding to the nonzero entries, fully describe the sparsity pattern of the considered tensor. In the context of CS with Gaussian measurement matrices, the Cramér–Rao bound (CRB) on the estimation accuracy of a Bernoulli-distributed block-sparse core tensor is derived. This prior assumes that each entry of the core tensor has a given probability to be nonzero, leading to random supports of truncated Binomial-distributed cardinalities. Based on the limit form of the Poisson distribution, an approximated CRB expression is given for large dictionaries and a highly block-sparse core tensor. Using the property that the mode unfolding matrices of a block-sparse tensor follow the multiple-measurement vectors (MMV) model with a joint sparsity pattern, a fast and accurate estimation scheme, called Beamformed mOde based Sparse Estimator (BOSE), is proposed in the second part of this paper. The main contribution of the BOSE is to "map" the MMV model onto the single MV model, thanks to beamforming techniques. Finally, the proposed performance bounds and the BOSE are applied in the context of CS to 1) nonbandlimited multidimensional signals with separable sampling kernels and 2) for multipath channels in a multiple-input multiple-output wireless communication scheme.**

***Index Terms*—Compressed sensing, sparse tensor, beamforming, multiple-measurement vectors, sampling of nonbandlimited multidimensional signal.**

I. Introduction

TENSORS or multi-way arrays are now a prominent research area in signal processing [1]–[6] and in particular at the era of big data processing [7]–[9]. However, structured tensors represent a relatively little-examined topic. Ignoring this *a priori* knowledge leads to sub-optimal algorithms and degraded estimation performances. We can find several types of structured tensors. The first type of structure occurs when the entries of the measured tensor follow a particular arrangement as Toeplitz or Hankel, for instance, in case of higher-order statistics tensors [10]–[14] or symmetric in the decomposition of high-order Volterra series [15]. The second type of structure arises in the decomposition of a measured tensor into the q-mode products of an unstructured core tensor with a sequence of structured factor matrices [16]. The last type of structure occurs when the core tensor admits a specific structure. The most evident case occurs when the core tensor is diagonal as in the Multidimensional Harmonic Retrieval problem [17]–[19], but a different structure appears when only a few entries of the core tensor are non-zeros, *i.e.*, the core tensor is sparse.

The compressed sensing theory [20]–[22] is now a very well established framework. Compressive Sampling (CS) is closely related to the compressed sensing philosophy. In CS, the data acquisition is performed at a rate lower than the Shannon sampling rate [23]. CS has been extensively studied from a theoretical point of view [24], [25] and has been exploited in many operational applications as, for instance, in array processing [26], wireless communication [27], video processing [28] or in MIMO radar [29]. Recently, the natural complementarity of CS and sparse tensors has been demonstrated in [30]–[33] and the concept of block-sparse core tensors turns out to be a natural multidimensional extension of the notion of block-sparse vectors [34], [35] or also the Multiple-Measurement Vectors (MMV) model with joint sparsity pattern [36]–[39]. Block-sparse tensors arise naturally in a wide range of applications as, for instance, in Magnetic Resonance Imaging (MRI), hyper-spectral imaging, multidimensional inpainting, missing data problems for EEG, super-resolution imaging (see [32] and the references therein) or in MIMO wireless channel communication [39]. Our context is based on the definition of block-sparse tensors introduced by Caiafa and Cichocki [32]. In addition, our contribution follows a different "philosophy" to [31], [30] since the main goal in these works is to reduce the initial problem into parallelizable sub-problems for tensor mode recovery. Reference [33] studies restricted isometry/incoherence properties of the mode loading matrices; but it does not deal with tensor compression.

Performance bounds constitute an important research domain since they can predict the best estimation accuracy in terms of the Mean Square Error (MSE) [40], [41]. In the CS context, performance bounds for the estimation accuracy of sparse vectors have been derived in [42]–[47], for instance. In the tensor

Manuscript received October 21, 2015; revised June 3, 2016; accepted July 23, 2016. Date of publication August 16, 2016; date of current version October 4, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yimin D. Zhang. This work was supported by the following projects: MAGELLAN (ANR-14-CE23-0004-01), MI-CNRS TITAN, MH ICode blanc, and the DFG project EXPRESS (HA 2239/6-1).

R. Boyer is with the Laboratoire des Signaux et Systèmes, University of Paris-Sud, Gif-Sur-Yvette 91190, France (e-mail: remy.boyer@l2s.centralesupelec.fr).

M. Haardt is with the Communications Research Laboratory, Ilmenau University of Technology, Ilmenau D-98684, Germany (e-mail: Martin.Haardt@tu-ilmenau.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2016.2600510

context, performance bounds have been proposed in [16], [19], [48], [49]. The goal of this work is to study the estimation performance of the non-zero entries of the block sparse core tensor from noisy compressive measurements. Unlike the existing approaches, the proposed analysis is performed (i) in the case of random supports with random cardinalities, *i.e.*, each entry of the block-sparse core tensor has a given probability to take a non-zero value and (ii) for multidimensional Gaussian measurements. Note that the assumption (i) models our uncertainty knowledge on the support sets and on the cardinalities. Many popular sparsity based estimators as for instance the Orthogonal Matching Pursuit (OMP) [50], [51], the Compressive Sampling Matching Pursuit (CoSaMP) [52], the Basis Pursuit (BP), the BP De-Noising (BPDN) [53], the Lasso [54] or again the Iterative Hard Thresholding (IHT) [55] algorithms have been historically developed to solve the SMV model. Based on this fact, a second goal of this work is to propose an efficient and generic pre-processing strategy to adapt any sparsity based estimators to the MMV model. Our results are illustrated for CS of non-bandlimited multidimensional signals [56]–[58] with separable sampling kernels and for compressive channel estimation for wireless MIMO systems.

This article contains four main sections. The first part introduces the noisy compressive sampling context with block-sparse tensors. In particular, the structure of the vectorized measurement tensor and the mode unfolding matrices are described. The second part presents a closed-form expression of the CRB for the estimation of the non-zero entries of the core tensor with a random support of random cardinalities. The third part proposes a new fast estimator called BOSE for a Beamformed mOde Sparse Estimator adapted to the estimation of the entries of a block-sparse core tensor. Finally, the proposed lower bound and the BOSE are applied in the context of CS of non-bandlimited multidimensional signals with separable sampling kernels.

Notation: The notation used through this paper is the following: scalars, vectors, matrices and tensors are represented by italic lower-case, boldface lower-case, boldface upper-case and boldface calligraphic upper-case symbols, respectively. Sets are denoted by calligraphic upper-case symbols, *e.g.*, $\mathcal{X}$ and its i-th element is $\mathcal{X}\{i\}$. The union of two sets is denoted by $\cup$. The symbols $(\cdot)^T$, $(\cdot)^\dagger$, $\mathrm{Tr}(\cdot)$ and $(\cdot)!$ denote the transpose, the pseudo-inverse, the trace operator and the factorial, respectively. Furthermore, $\mathrm{Binomial}(N, P)$ stands for the Binomial distribution parametrized by N independent yes/no experiments, with a success probability P [59]. The Binomial coefficient is defined by $\binom{a}{b} = \frac{a!}{b!(a-b)!}$. Moreover, $\mathcal{N}(\mu, \sigma^2)$ stands for the real Gaussian probability density function (pdf) with mean μ and variance σ^2, and Poisson (θ) stands for the Poisson distribution with parameter θ. The distributions χ^2_n and $\chi^2_n(v)$ stand for the central and the non-central chi-squared distribution with n degrees of freedom and non-centrality parameter v. In addition, the function $Q_{\mathcal{X}}(\cdot)$ denotes the right tail of the distribution $\mathcal{X}$. The notation $\sim$ (*resp.* $\overset{a}{\sim}$) means that a random variable follows (*resp.* asymptotically follows) a particular distribution. The symbol $\mathrm{diag}(\mathbf{x})$ denotes a diagonal matrix, where the elements of the vector $\mathbf{x}$ specify its diagonal elements. Furthermore, $1_{\mathcal{X}}(x)$ is the indicator function with respect to the set $\mathcal{X}$, *i.e.*, $1_{\mathcal{X}}(x) = 1$ if $x \in \mathcal{X}$ and 0 otherwise. The symbol $\bullet$, denotes the Schur-Hadamard product. Moreover, $\delta(\cdot, \ldots, \cdot)$ is the multidimensional Dirac delta symbol and $\bar{\delta}(\cdot)$ is the Kronecker delta. The symbol $\mathrm{vec}(\cdot)$ stands for the vectorization operator which converts the $S \times T$ matrix $\mathbf{X} = [\mathbf{x}_1 \ldots \mathbf{x}_T]$ into a $(ST) \times 1$ vector $\mathbf{x}$ obtained by staking the T columns of the matrix $\mathbf{X}$ according to $\mathbf{x} = [\mathbf{x}_1^T \ldots \mathbf{x}_T^T]^T$. The matrix $\mathrm{unfold}_q(\boldsymbol{\mathcal{X}})$ is the q-mode unfolding of the tensor $\boldsymbol{\mathcal{X}}$ [6], i.e., the matrix of all q-mode vectors where the q-th index is varied in each column, and all other indices are kept fixed. In this paper, we use the forward column ordering, where we start by varying the first index, then the second, up to index $(q-1)$, continue with index $(q+1)$ up to index Q. The scalar entry localized at indexes $m_1, \ldots, m_Q$ in the tensor $\boldsymbol{\mathcal{X}}$ is denoted by $[\boldsymbol{\mathcal{X}}]_{m_1,\ldots,m_Q}$. The tensor, denoted by $[\boldsymbol{\mathcal{X}}]_{\mathcal{X},\mathcal{Y},\mathcal{Z}}$ and extracted from the larger tensor $\boldsymbol{\mathcal{X}}$ contains the entries $[\boldsymbol{\mathcal{X}}]_{x,y,z}$ with $\forall(x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, where $\times$ denotes the cartesian product. Furthermore, $\mathcal{R}(\cdot)$ stands for the range space. The notation $\lambda_k(\mathbf{X})$ denotes the k-th eigenvalue of the positive semi-definite matrix $\mathbf{X}$. Moreover, $\lambda_{\max}(\mathbf{X})$ is defined as the maximal eigenvalue of the eigenspectrum of $\mathbf{X}$, *i.e.*, $\lambda_{\max}(\mathbf{X}) = \max_k \lambda_k(\mathbf{X})$. The operator $\mathcal{P}_{\max}(\mathbf{X})$ computes the eigenvector associated with the largest eigenvalue $\lambda_{\max}(\mathbf{X})$ of the matrix $\mathbf{X}$. Furthermore, $\det(\cdot)$ denotes the determinant. The (r, p)-norm is defined according to $||\mathbf{X}||_{r,p} = (\sum_{i=1}^N ||[\mathbf{X}]_i||_r^p)^{1/p}$. Finally, $O(\cdot)$ stands for the big-O notation.

II. Noisy Compressive Acquisition With Block-Sparse Tensors

A. The Compressed Sensing Framework

1) Single Measurement Vector (SMV): Let $\mathbf{y}$ be the $N \times 1$ noisy measurement vector in a (standard) compressed sensing (CS) model [21], [24]:

$$\mathbf{z} = \boldsymbol{\Psi}\mathbf{r} + \mathbf{w}, \tag{1}$$

where $\mathbf{w}$ represents a zero-mean white Gaussian noise of unknown variance σ^2 and $\boldsymbol{\Psi}$ is the $N \times K$ measurement matrix. Let $\mathbf{r} \overset{\text{def.}}{=} \boldsymbol{\Phi}\mathbf{x}^{\text{spar}}$ where the matrix $\boldsymbol{\Phi}$ is a $K \times K$ orthonormal basis and $\mathbf{x}^{\text{spar}} \in \mathcal{W}_M \overset{\text{def.}}{=} \{\mathbf{x}^{\text{spar}} \in \mathbb{R}^K, \|\mathbf{x}^{\text{spar}}\|_0 = \mathrm{card}(\mathcal{S}) \leq M\}$ with $\|\cdot\|_0$ denoting the pseudo-norm l_0 and $\mathcal{S}$ being the support of $\mathbf{x}^{\text{spar}}$, *i.e.*, the collection of the positions of the non-zero entries in vector $\mathbf{x}^{\text{spar}}$. Defining the $N \times K$ dictionary matrix $\mathbf{D} \overset{\text{def.}}{=} \boldsymbol{\Psi}\boldsymbol{\Phi}$, the model in eq. (1) can be recast as

$$\mathbf{z} = \mathbf{D}\mathbf{x}^{\text{spar}} + \mathbf{w}. \tag{2}$$

The optimization problem is given by

$$\min ||\mathbf{x}^{\text{spar}}||_0 \text{ subject to } ||\mathbf{z} - \mathbf{D}\mathbf{x}^{\text{spar}}|| \leq \epsilon$$

for a positive ϵ. The above optimization problem is NP-hard because of the nonconvexity of the l_0 pseudo-norm. Furthermore, it is standard to relax the above problem by considering the l_1 norm of $\mathbf{x}^{\text{spar}}$ instead of the pseudo-norm l_0. By doing this, the relaxed optimization problem is convex, and thus can be implemented as a linear program.

2) Multiple Measurement Vectors (MMV): The MMV model is an extension of the SMV model given by eq. (2) for $\{\mathbf{z}(t),\ 1 \leq t \leq T\}$ where T is the number of snapshots. Generalizing the model of eq. (2), we obtain

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}(1) & \dots & \mathbf{z}(T) \end{bmatrix} = \mathbf{D}\mathbf{X}^{\text{spar}} + \mathbf{W}$$

where $\mathbf{W} = \begin{bmatrix} \mathbf{w}(1) & \dots & \mathbf{w}(T) \end{bmatrix}$ is the matrix composed by multiple noise vectors and

$$\mathbf{X}^{\text{spar}} = \begin{bmatrix} \mathbf{x}^{\text{spar}}(1) & \dots & \mathbf{x}^{\text{spar}}(T) \end{bmatrix}.$$

Note that the matrix $\mathbf{X}^{\text{spar}}$ consists of a set of jointly sparse vectors with a common support [38], [60]. The MMV model aims to recover the sparse representations of SMVs simultaneously. The MMV model is useful in several operational contexts as for instance for recovery of sparse brain excitations [36], in direction of arrival (DOA) estimation [61] or in wireless communication [39]. The convex relaxed optimization problem for the MMV model is given by

$$\min ||\mathbf{X}^{\text{spar}}||_{2,1} \text{ subject to } ||\mathbf{Z} - \mathbf{D}\mathbf{X}^{\text{spar}}|| \leq \epsilon.$$

3) Pseudo-Isometric Random Dictionary: Classical sampling theory says that, to ensure that no information is lost, the number of measurements, N, should be at least equal to K. In contrast, in CS theory this goal is reached for $N \ll K$ as long as the $K \times 1$ amplitude vector $\mathbf{x}^{\text{spar}}$ is sparse in a given basis $\mathbf{\Phi}$ (*e.g.*, the canonical basis of $\mathbb{R}^K$, the Fourier basis, etc.) [20]. This allows one to consider the CS theory to solve the ill-posed problem where the dictionary $\mathbf{D}$ is a redundant matrix. A fundamental question in CS is to determine how large N must be to enable the recovery of $\mathbf{x}^{\text{spar}}$.

Remark 2.1: [62] Assume that $\mathbf{D}$ is drawn according to a distribution for which the concentration inequality [21], [22], [63] holds for $\delta \in (0, 1)$:

$$\Pr\left(\left| \|\mathbf{D}\mathbf{x}^{\text{spar}}\|^2 - \|\mathbf{x}^{\text{spar}}\|^2 \right| \geq \delta \|\mathbf{x}^{\text{spar}}\|^2\right) \leq 2e^{-\delta^2 N} \quad (3)$$

then all M-sparse vectors can be stably recovered with high probability from a number of measurement given by $N = O\left(M \log \frac{K}{M}\right) < K$ [64].

4) Universal Design Strategy: Determining whether the dictionary $\mathbf{D}$ satisfies the the concentration inequality is combinatorially complex but a strategy called universal has been introduced, for instance, in [21], [22], [63]. Assume that $\mathbf{\Psi}$ is an orthonormal basis and $\mathbf{\Phi}$ is drawn from independent and identically distributed Gaussian entries of zero mean and variance $1/N$. Recall that Gaussian matrices satisfy the concentration inequality [62]. Consequently, the adopted strategy leads to a dictionary $\mathbf{D}$ fulfilling Remark 3.1.

B. Extension to Block-Sparse Tensors

1) Preliminary Notions:

Definition 2.2: The Kronecker product of matrices $\mathbf{X}$ and $\mathbf{Y}$ of size $I \times J$ and $K \times N$, respectively is given by

$$\mathbf{X} \otimes \mathbf{Y} = \begin{bmatrix} [\mathbf{X}]_{11}\mathbf{Y} & \dots & [\mathbf{X}]_{1J}\mathbf{Y} \\ \vdots & & \vdots \\ [\mathbf{X}]_{I1}\mathbf{Y} & \dots & [\mathbf{X}]_{IJ}\mathbf{Y} \end{bmatrix} \in \mathbb{R}^{(IK)\times(JN)}.$$

Lemma 2.3 ([65]):

1) For non-singular matrices $\mathbf{X}$ and $\mathbf{Y}$, the following property holds $(\mathbf{X} \otimes \mathbf{Y})^{-1} = \mathbf{X}^{-1} \otimes \mathbf{Y}^{-1}$. Note that the non-singularity of the matrices $\mathbf{X}$ and $\mathbf{Y}$ means that $\mathbf{X} \otimes \mathbf{Y}$ is also non-singular.
2) In addition, we have $\text{Tr}(\mathbf{X} \otimes \mathbf{Y}) = \text{Tr}\mathbf{X} \cdot \text{Tr}\mathbf{Y}$.

Definition 2.4: The n-mode product denoted by $\times_n$ between a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{M_1 \times \ldots \times M_N}$ and a matrix $\mathbf{U} \in \mathbb{R}^{K \times M_n}$ is denoted by $\boldsymbol{\mathcal{X}} \times_n \mathbf{U} \in \mathbb{R}^{M_1 \times \ldots \times M_{n-1} \times K \times M_{n+1} \times \ldots \times M_N}$ with

$$\begin{aligned}[\boldsymbol{\mathcal{X}} \times_n \mathbf{U}]_{m_1,\ldots,m_{n-1},k,m_{n+1},\ldots,m_N} \\ = \sum_{m_n=1}^{M_n} [\boldsymbol{\mathcal{X}}]_{m_1,\ldots,m_N} [\mathbf{U}]_{k,m_n}\end{aligned}$$

where $1 \leq k \leq K$.

Definition 2.5: [66] The Tucker model of order N is defined according to

$$[\boldsymbol{\mathcal{X}}]_{k_1,\ldots,k_N} = \sum_{m_1,\ldots,m_N} [\boldsymbol{\mathcal{S}}]_{m_1,\ldots,m_N} \prod_{n=1}^{N} [\mathbf{U}_n]_{k_n,m_n}$$

where $\boldsymbol{\mathcal{S}}$ is the $M_1 \times \ldots \times M_N$ core tensor and $\mathbf{U}_n$ is the n-th factor matrix of size $K_n \times M_n$. An equivalent formulation using the n-mode product is

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{S}} \times_1 \mathbf{U}_1 \times_2 \ldots \times_N \mathbf{U}_N \in \mathbb{R}^{K_1 \times \ldots \times K_N}.$$

Definition 2.6: The n-mode unfolding matrix of size $M_n \times \left(\prod_{k=1,k\neq n}^{N} M_k\right)$ denoted by $\mathbf{X}^{(n)} = \text{unfold}_n(\boldsymbol{\mathcal{X}})$ of a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{M_1 \times \ldots \times M_N}$ is defined according to

$$[\mathbf{X}^{(n)}]_{m_n,h} = [\boldsymbol{\mathcal{X}}]_{m_1,\ldots,m_N}$$

where $h = 1 + \sum_{k=1,k\neq n}^{N} (m_k - 1) \prod_{v=1,v\neq n}^{k-1} M_v$. The n-mode unfolding admits the following decomposition:

$$\mathbf{X}^{(n)} = \mathbf{U}_n \mathbf{S}^{(n)} \left(\mathbf{U}_N \otimes \ldots \otimes \mathbf{U}_{n+1} \otimes \mathbf{U}_{n-1} \ldots \otimes \mathbf{U}_1\right)^T \quad (4)$$

where $\mathbf{S}^{(n)} = \text{unfold}_n(\boldsymbol{\mathcal{S}})$ is n-mode unfolding of the core tensor.

Definition 2.7: Let $\mathbf{x}$ be the $\left(\prod_{n=1}^{N} M_n\right) \times 1$ vectorization representation of the tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{M_1 \times \ldots \times M_N}$. The vector $\mathbf{x}$ is defined as the columns stacking of the 1-mode unfolding matrix, *i.e.*, $\mathbf{x} = \text{vec}\mathbf{X}^{(1)}$. Using eq. (4) and the property $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}\mathbf{B}$, we get

$$\mathbf{x} = (\mathbf{U}_N \otimes \ldots \otimes \mathbf{U}_2) \otimes \mathbf{U}_1 \text{vec}\mathbf{S}^{(1)} = \mathbf{U}\mathbf{s}$$

where $\mathbf{s} = \text{vec}\mathbf{S}^{(1)}$ and $\mathbf{U} = \mathbf{U}_N \otimes \ldots \otimes \mathbf{U}_2 \otimes \mathbf{U}_1$.

2) Definition of a Block-Sparse Tensor: Assume that the real-valued tensor $\boldsymbol{\mathcal{R}}$ of size $K_1 \times \ldots \times K_Q$ follows a Tucker model (see Definition 2.5):

$$\boldsymbol{\mathcal{R}} = \boldsymbol{\mathcal{X}}^{\text{spar}} \times_1 \mathbf{\Phi}_1 \times_2 \ldots \times_Q \mathbf{\Phi}_Q \quad (5)$$

where the $K_1 \times \ldots \times K_Q$ tensor $\boldsymbol{\mathcal{X}}^{\text{spar}}$ is $(M_1 \times \ldots \times M_Q)$-block sparse with respect to a set of Q basis matrices $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_Q$ each of size $K_q \times K_q$ with $M_q \ll K_q$. More formally, define the q-th support set $\mathcal{F}_q$ composed by the M_q indices corresponding to the non-zero values in the q-th dimension of the tensor $\boldsymbol{\mathcal{X}}^{\text{spar}}$. Let $\mathcal{F} = \mathcal{F}_1 \times \ldots \times \mathcal{F}_Q$ be the cartesian product set associated with the Q supports sets, then the block-sparsity property is defined by

$$[\boldsymbol{\mathcal{X}}^{\mathrm{spar}}]_{m_1,\ldots,m_Q} \begin{cases} \neq 0, & \forall (m_1,\ldots,m_Q) \in \mathcal{F}, \\ = 0, & \text{otherwise.} \end{cases}$$

In the context of the CS, Q measurement/sensing matrices, denoted by $\boldsymbol{\Psi}_1,\ldots,\boldsymbol{\Psi}_Q$, each of size $I_q \times K_q$ with $I_q < K_q$ are introduced to achieve dimensionality reduction of the available measurement tensor according to

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{R}} \times_1 \boldsymbol{\Psi}_1 \times_2 \ldots \times_Q \boldsymbol{\Psi}_Q. \tag{6}$$

Now, introduce Q overcomplete dictionaries $\mathbf{D}_1,\ldots,\mathbf{D}_Q$ with $\mathbf{D}_q = \boldsymbol{\Psi}_q \boldsymbol{\Phi}_q$. Using eq. (5) and eq. (6), the noisy CS model is given by

$$\begin{aligned} \boldsymbol{\mathcal{Z}} &= \boldsymbol{\mathcal{X}}^{\mathrm{spar}} \times_1 \boldsymbol{\Phi}_1 \times_2 \ldots \times_Q \boldsymbol{\Phi}_Q \\ &\quad \times_1 \boldsymbol{\Psi}_1 \times_2 \ldots \times_Q \boldsymbol{\Psi}_Q + \boldsymbol{\mathcal{W}} \\ &= \boldsymbol{\mathcal{X}}^{\mathrm{spar}} \times_1 \mathbf{D}_1 \times_2 \ldots \times_Q \mathbf{D}_Q + \boldsymbol{\mathcal{W}} \end{aligned} \tag{7}$$

in which each entry of the noise tensor $\boldsymbol{\mathcal{W}}$ has additive, circular and Gaussian entries that are uncorrelated in the Q dimensions according to $[\boldsymbol{\mathcal{W}}]_{k_1,\ldots,k_Q} \sim \mathcal{N}(0, \sigma^2 \prod_{q=1}^{Q} \bar{\delta}(k_q - k_{q'}))$. A more compact and useful expression can be introduced when the zero values in $\boldsymbol{\mathcal{X}}^{\mathrm{spar}}$ are removed. Define the $I_q \times M_q$ matrix $\mathbf{D}_{\mathcal{F}_q} = \boldsymbol{\Psi}_q \boldsymbol{\Phi}_{\mathcal{F}_q}$ in which the $K_q \times M_q$ matrix $\boldsymbol{\Phi}_{\mathcal{F}_q}$ corresponds to the M_q columns of $\boldsymbol{\Phi}_q$ associated with the support $\mathcal{F}_q$. Define $\boldsymbol{\mathcal{X}}_{\mathcal{F}}$ as the $M_1 \times \ldots \times M_Q$ tensor constituted by the non-zero entries in $\boldsymbol{\mathcal{X}}^{\mathrm{spar}}$. It is now straightforward to obtain an alternative and more compact expression of the model

$$\boldsymbol{\mathcal{Z}} = \boldsymbol{\mathcal{X}}_{\mathcal{F}} \times_1 \mathbf{D}_{\mathcal{F}_1} \times_2 \ldots \times_Q \mathbf{D}_{\mathcal{F}_Q} + \boldsymbol{\mathcal{W}}, \tag{8}$$

in which the block-sparsity property can be written in the following manner:

$$\boldsymbol{\mathcal{X}}_{\mathcal{F}} = [\boldsymbol{\mathcal{X}}^{\mathrm{spar}}]_{\mathcal{F}_1,\ldots,\mathcal{F}_Q}. \tag{9}$$

C. MMV Structured q-Mode Unfoldings

Let $\mathbf{Y}^{(q)} = \mathrm{unfold}_q(\boldsymbol{\mathcal{Y}})$ and $\mathbf{W}^{(q)} = \mathrm{unfold}_q(\boldsymbol{\mathcal{W}})$ be the q-mode unfoldings given in Definition 2.6, *i.e.*, the matrix unfolding along the q-th dimension, of the tensors $\boldsymbol{\mathcal{Y}}$ and $\boldsymbol{\mathcal{W}}$, respectively. Using the mode decomposition specified in Definition 2.6, the noisy q-mode unfolding of the tensor $\boldsymbol{\mathcal{Z}}$ is given by

$$\mathbf{Z}^{(q)} = \mathrm{unfold}_q(\boldsymbol{\mathcal{Z}}) = \mathbf{Y}^{(q)} + \mathbf{W}^{(q)} = \mathbf{D}_q \mathbf{Q}_q + \mathbf{W}^{(q)}, \tag{10}$$

where $\mathbf{Z}^{(q)}$ is a $I_q \times \tilde{I}_q$ matrix with $\tilde{I}_q = \bar{I}/I_q$ and

$$\begin{aligned} \mathbf{Q}_q &= \mathbf{X}^{(q)}_{\mathrm{spar}} \left(\mathbf{D}_Q \otimes \ldots \otimes \mathbf{D}_{q+1} \otimes \mathbf{D}_{q-1} \ldots \otimes \mathbf{D}_1\right)^T \\ &= \mathbf{X}^{(q)}_{\mathcal{F}} \left(\mathbf{D}_{\mathcal{F}_Q} \otimes \ldots \otimes \mathbf{D}_{\mathcal{F}_{q+1}} \otimes \mathbf{D}_{\mathcal{F}_{q-1}} \ldots \otimes \mathbf{D}_{\mathcal{F}_1}\right)^T \end{aligned} \tag{11}$$

in which $\mathbf{X}^{(q)}_{\mathrm{spar}} = \mathrm{unfold}_q(\boldsymbol{\mathcal{X}}^{\mathrm{spar}})$ and $\mathbf{X}^{(q)}_{\mathcal{F}} = \mathrm{unfold}_q(\boldsymbol{\mathcal{X}}_{\mathcal{F}})$. This particular structure in the q-mode unfoldings leads to the following remark.

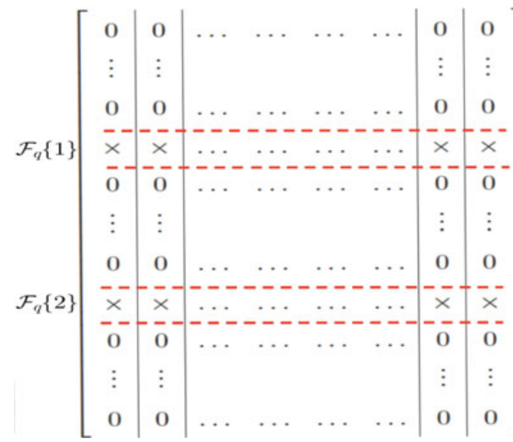

Fig. 1. Sparsity structure of the matrix $\mathbf{Q}_q$ for the support $\mathcal{F}_q = \{\mathcal{F}_q\{1\}, \mathcal{F}_q\{2\}\}$ of cardinality $M_q = 2$. The symbol $\times$ stands for a non-zero entry. Clearly, each columns exhibits the same sparsity pattern.

Remark 2.8: First note that thanks to the block-sparse property assumption of the core tensor, the columns of the matrix $\mathbf{Q}_q$ are M_q-sparse with a common support $\mathcal{F}_q$ (see Fig. 1). This structure is well-known under the name of Multiple-Measurement Vectors (MMV) model with a joint sparsity pattern.

D. Kronecker Structured Dictionary

Using Definition 2.7, the vectorized representation of the tensor $\boldsymbol{\mathcal{Z}}$ given in eq. (7) is

$$\mathbf{z} = \mathrm{vec}\mathbf{Z}^{(1)} = \mathbf{D}\mathbf{x}^{\mathrm{spar}} + \mathbf{w} \tag{12}$$

where $\mathbf{w} = \mathrm{vec}\mathbf{W}^{(1)} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, $\mathbf{x}^{\mathrm{spar}} = \mathrm{vec}\mathbf{X}^{(1)}$ is a sparse vector, and $\mathbf{D} = \mathbf{D}_Q \times \ldots \times \mathbf{D}_1$. Given the knowledge of the mapping function between the supports $\mathcal{S}$ and $\mathcal{F}$, the following relation holds

$$\mathbf{D}\mathbf{x}^{\mathrm{spar}} = \mathbf{D}_{\mathcal{F}} \mathbf{x}_{\mathcal{S}},$$

where

$$\mathbf{D}_{\mathcal{F}} = \mathbf{D}_{\mathcal{F}_Q} \otimes \ldots \otimes \mathbf{D}_{\mathcal{F}_1}$$

and

$$\mathbf{x}_{\mathcal{S}} = [\mathbf{x}^{\mathrm{spar}}]_{\mathcal{S}} = \mathrm{vec}\left(\mathrm{unfold}_1(\boldsymbol{\mathcal{X}}_{\mathcal{F}})\right) \tag{13}$$

are a dense (*i.e.*, all entries are non-zero) tensor and vector, respectively. This expression characterizes the block-sparse property and is an alternative formulation of eq. (9) in a vectorial formalism. The $(1 + \sum_{q=1}^{Q} (m_q - 1) \prod_{k=1,k\neq q}^{q-1} M_k)$-th entry of the support $\mathcal{S}$ is given by

$$1 + \sum_{q=1}^{Q} \left(\mathcal{F}_q\left\{m_q\right\} - 1\right) \prod_{k=1,k\neq q}^{q-1} K_k \tag{14}$$

for $1 \leq m_q \leq M_q$.

Example 1: The vectorization of a third-order $(Q = 3)$ (M_1, M_2, M_3)-block sparse tensor is illustrated in Fig. 2.The entry u in $\boldsymbol{\mathcal{X}}^{\mathbf{spar}}$ is located at

$$\Big(\mathcal{F}_1\{m_1 = 2\} = 3, \mathcal{F}_2\{m_2 = 1\} = 2, \mathcal{F}_3\{m_3 = 2\} = 2\Big).$$

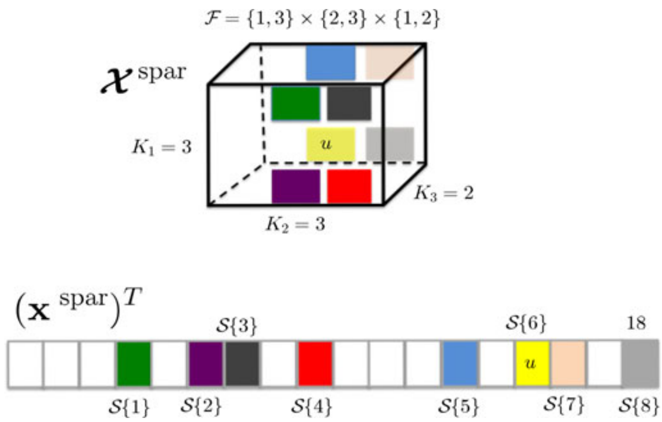

Fig. 2. Illustrative example of the vectorization of a third-order (M_1, M_2, M_3)-block sparse tensor with $M_1 = M_2 = M_3 = 2$.

According to eq. (14), the $(m_1 + (m_2 - 1)M_1 + (m_3 - 1)M_1 M_2)$-th entry of support $\mathcal{S}$, is given by

$$\mathcal{F}_1\{2\} + (\mathcal{F}_2\{1\} - 1)K_1 + (\mathcal{F}_3\{2\} - 1)K_1 K_2.$$

Finally, we get $\mathcal{S}\{6\} = 15$.

III. Performance Analysis for Random Supports of Random Cardinalities

A. Stable Recovery Guarantee

CS with Kronecker structured dictionaries has been extensively studied in for instance [32], [67]–[69]. In the following, Remark 3.1 provides an extension to the multidimensional case.

Remark 3.1: [62] Assume that $\mathbf{D} = \mathbf{D}_Q \otimes \ldots \otimes \mathbf{D}_1$ is drawn according to a pdf for which the concentration inequality given by eq. (3) holds for $\delta \in (0, 1)$ where $N = \prod_{q=1}^{Q} N_q$. Then, all $(M_1, \ldots, M_Q)$-sparse vectors can be stably recovered with high probability from a number of measurement given by

$$N = O\left(\prod_{q=1}^{Q} M_q \sum_{q=1}^{Q} \log \frac{K_q}{M_q}\right).$$

B. Description of the Statistical Prior for the Core Tensor

A natural model to introduce a random activation/selection mechanism of the entries of a $K_1 \times \ldots \times K_Q$ dense (non-sparse) tensor, denoted by $\boldsymbol{\mathcal{X}}$, is described in this section. Let $\boldsymbol{\mathcal{X}}^{\text{spar}}$ be a $(M_1, \ldots, M_Q)$-block sparse tensor on the supports $\{\mathcal{F}_1, \ldots, \mathcal{F}_Q\}$. The tensor $\boldsymbol{\mathcal{X}}^{\text{spar}}$ is defined according to

$$\boldsymbol{\mathcal{X}}^{\text{spar}} = \boldsymbol{\mathcal{X}} \bullet \boldsymbol{\mathcal{Q}} \tag{15}$$

where the binary-valued entries of the $K_1 \times \ldots \times K_Q$ tensor $\boldsymbol{\mathcal{Q}}$ activate or not the corresponding entries of the core tensor $\boldsymbol{\mathcal{X}}$ according to the relation $[\boldsymbol{\mathcal{Q}}]_{m_1,\ldots,m_Q} = \prod_{q=1}^{Q} 1_{\mathcal{F}_q}(m_q)$. Using Definition 2.4 of the q-mode product, eq. (15) can be reformulated according to $\boldsymbol{\mathcal{X}}^{\text{spar}} = \boldsymbol{\mathcal{X}} \times_1 \mathbf{V}_{\mathcal{F}_1} \times_2 \ldots \times_Q \mathbf{V}_{\mathcal{F}_Q}$ in which $\mathbf{V}_{\mathcal{F}_q} = \text{diag}\{1_{\mathcal{F}_q}(1), \ldots, 1_{\mathcal{F}_q}(K_q)\}$ is a $K_q \times K_q$ diagonal selection matrix. This matrix has to

1) guarantee the model identifiability constraint with high probability

$$P_q^{\text{id}} = \Pr\Big(M_q \in \mathcal{V}_q = \{1, \ldots, I_q - 1\}\Big), \tag{16}$$

i.e., for all q, the number of unknown parameters, M_q, does not exceed the number of available measurements, I_q.

2) introduce a statistical prior which randomly "activates or not" the entries of the core tensor. Towards this goal, define the probability of the event "$m_q \in \mathcal{F}_q$", *i.e.*,

$$P_q = \Pr(m_q \in \mathcal{F}_q)$$

and define $1_{\mathcal{F}_1}(m_1), \ldots, 1_{\mathcal{F}_Q}(m_Q)$ as Q mutually independent Bernoulli random variables such that $\Pr(1_{\mathcal{F}_q}(m_q) = 1) = P_q$ and $\Pr(1_{\mathcal{F}_q}(m_q) = 0) = 1 - P_q$. Bernoulli priors are widely used in the literature (see [46], [70]–[75] for instance).

The cardinality of $\mathcal{F}_q$ is the number of "1" s on the diagonal of the matrix $\mathbf{V}_{\mathcal{F}_q}$, *i.e.*, $M_q = \text{Tr}\mathbf{V}_{\mathcal{F}_q} = \sum_{m_q=1}^{K_q} 1_{\mathcal{F}_q}(m_q)$. So, M_q is the sum of K_q mutually independent Bernoulli random variables and thus $M_q \sim \text{Binomial}(K_q, P_q)$ with $\mathbb{E}M_q = K_q P_q = \bar{M}_q$, where $\bar{M}_q$ is the *a priori* fixed mean cardinality [59]. As the cardinalities $\{M_1, \ldots, M_Q\}$ are generally unknown, it makes sense to assume that these parameters are i.i.d. random.

Observe that the assumed statistical model in point 2) can potentially violate the model identifiability constraint described in the point 1). So, introduce the truncated Binomial distribution [59] on the interval $\mathcal{V}_q$ according to

$$\Pr\left(M_q = m_q \middle| M_q \in \mathcal{V}_q\right) = \frac{\Pr(M_q = m_q) 1_{\mathcal{V}_q}(m_q)}{P_q^{\text{id}}}$$

where $P_q^{\text{id}} = \sum_{m_q \in \mathcal{V}_q} \binom{K_q}{m_q} P_q^{m_q} (1 - P_q)^{K_q - m_q}$.

C. CRB for Random Supports With Truncated Binomial-Distributed Cardinalities

Let $\mathcal{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_Q\}$ where $\mathcal{M}_q = \{M_q | M_q \in \mathcal{V}_q\}$. Define an unbiased estimated vector $\hat{\mathbf{x}}_{\mathcal{S}}(\mathbf{z}, \mathbf{D}_{\mathcal{F}}, \mathcal{M})$ of $\mathbf{x}_{\mathcal{S}}$ as in eq. (13). The averaged MSE is given by

$$\text{MSE}_{av.} = \mathbb{E}_{\mathcal{M}} \mathbb{E}_{\mathbf{D}_{\mathcal{F}}|\mathcal{M}} \text{MSE} \tag{17}$$

where

$$\text{MSE} = \frac{1}{\bar{I}} \mathbb{E}_{\mathbf{z}|\mathbf{D}_{\mathcal{F}},\mathcal{M}} \left\| \mathbf{x}_{\mathcal{S}} - \hat{\mathbf{x}}_{\mathcal{S}}(\mathbf{z}, \mathbf{D}_{\mathcal{F}}, \mathcal{M}) \right\|^2$$

in which $\bar{I} = \prod_{q=1}^{Q} I_q$.

Result 3.2: Assume that the noisy real-valued measured tensor follows the model of eq. (8), where the unknown core tensor is $(M_1, \ldots, M_Q)$-block sparse. Consider the following assumptions:

1) We are given Q random overcomplete dictionaries $\mathbf{D}_1, \ldots, \mathbf{D}_Q$ defined according to the universal design strategy of Section II-A4.
2) Moreover, we assume the statistical model introduced in Section III-A and the known mean cardinalities $\{\bar{M}_1, \ldots, \bar{M}_Q\}$.

In this scenario, the averaged $\text{MSE}_{av.}$ defined by eq. (17) for the estimation of the non-zero entries of the core tensor is lower bounded by the following expected CRB:

$$\mathcal{C} = \sigma^2 \prod_{q=1}^{Q} \frac{1}{P_q^{\text{id}}} \sum_{m_q \in \mathcal{V}_q} \frac{m_q}{I_q - m_q} \binom{K_q}{m_q} P_q^{m_q} (1 - P_q)^{K_q - m_q}. \tag{18}$$

Proof: See the Appendix A. ■

D. Approximated CRB for Large Dictionaries and a Highly Sparse Core Tensor

We are interested in deriving a bound $\mathcal{C}$ for

- largely redundant dictionaries, *i.e.*, $\forall q,\ K_q \to \infty$, and
- a highly $(\bar{M}_1, \ldots, \bar{M}_Q)$-block sparse core tensor, *i.e.*, for sufficiently small P_q such that $K_q P_q = \bar{M}_q$ is finite.

The corresponding bound is defined by

$$\mathcal{C}_\infty = \lim_{K_1,\ldots,K_Q \to \infty} \mathcal{C} \ \text{ s.t. } \lim_{K_q \to \infty} K_q P_q = \bar{M}_q < \infty$$

for $1 \le q \le Q$.

Unfortunately, the numerical computation of the lower bound given by eq. (18) is impractical for large K_q and thus there is a need to derive a closed-form expression of the bound $\mathcal{C}_\infty$. To this end, we use the limit form of the Poisson distribution (see [59], for instance), *i.e.*, for $K_q \to \infty$ and for sufficiently small P_q such that $\lim_{K_q \to \infty} K_q P_q = \bar{M}_q < \infty$, we have

$$M_q \overset{a}{\sim} \text{Poisson } (\bar{M}_q). \tag{19}$$

In the following remark, we specify the probability that the model identifiability constraint, denoted by P_q^{id} and defined in eq. (16), is fulfilled.

Remark 3.3:

1) The probability P_q^{id} for $K_q \to \infty$ can be approximated according to

$$\mathcal{P}_q^{\text{id}} = \lim_{K_q \to \infty} P_q^{\text{id}} \approx 1 - e^{-\bar{M}_q} - \epsilon(I_q), \tag{20}$$

where $\epsilon(I_q) = \lim_{K_q \to \infty} \Pr(I_q \le M_q \le K_q)$.

2) For $K_q, I_q \to \infty$, we have

$$\lim_{I_q \to \infty} \mathcal{P}_q^{\text{id}} = 1 - e^{-\bar{M}_q},$$

since $\epsilon(I_q) \to 0$.

Proof: The proof of the first statement is given in Appendix B. The proof of the second statement is straightforward and thus is omitted. ■

As we see the above probability can be considered as high because function $e^{-\bar{M}_q}$ is rapidly decreasing for growing $\bar{M}_q$. In the next result, an approximated CRB expression based on Result 3.2 is provided for large dictionaries and a highly sparse core tensor.

Result 3.4: For the scenario where $\forall q,\ \lim_{K_q \to \infty} K_q P_q = \bar{M}_q < \infty$ and $I_q \gg 1$, an approximated closed-form expression of the CRB given in Result 3.2 is

$$\mathcal{C}^\infty \approx \sigma^2 \prod_{q=1}^{Q} \frac{\bar{M}_q}{(1 - e^{-\bar{M}_q})(1 - e^{\bar{M}_q - I_q})(I_q - \bar{M}_q)}. \tag{21}$$

Proof: See the Appendix C. ■

Inspecting the bound given in eq. (21), we can note that for $\bar{M}_q \to I_q$, the terms in the denominator $1 - e^{\bar{M}_q - I_q}$ and $I_q - \bar{M}_q$ go to zero. This behavior is natural since in this case the degrees of freedom characterizing the system go to zero.

IV. Beamformed Mode Sparse Estimator (BOSE) for Block-Sparse Tensors

A. MMV to SMV Mapping - the Detection Theory Approach

Based on Remark 2.8, the set of sparse vectors sharing a common support is closed under any linear combination. This property has first been noticed in [61], [76] in the context of source localization for array processing and in ultrasound imaging [77]. Consequently, for any non-zero vector $\mathbf{h}_q$, $\mathbf{s}_q = \mathbf{Q}_q \mathbf{h}_q$ is a M_q-sparse vector on the support $\mathcal{F}_q$. In the context of the BOSE scheme, the vector $\mathbf{h}_q$ is viewed as a beamforming filter [77], [78].

Towards the derivation of an optimal beamforming filter, define the two-sided binary hypothesis detection problem [79]:

$$\begin{cases} \mathcal{H}_0 : \ [\mathbf{s}_q]_{\mathcal{F}_q} = \mathbf{0},\ \mathbf{z}_q \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{\mathbf{W}^{(q)}}) \\ \mathcal{H}_1 : \ [\mathbf{s}_q]_{\mathcal{F}_q} \neq \mathbf{0},\ \mathbf{z}_q \sim \mathcal{N}(\mathbf{D}_q \mathbf{s}_q, \mathbf{R}_{\mathbf{W}^{(q)}}). \end{cases} \tag{22}$$

Using the decorrelation of the noise with respect to the Q dimensions, we have $\hat{\mathbf{R}}_{\mathbf{W}^{(q)}} = \sigma_q^2 \mathbf{I}$ where $\sigma_q^2 = \sigma^2 ||\mathbf{h}_q||^2$.

A popular hypothesis discrimination criterion is the symmetrized Kullback-Leibler Divergence (KLD) [80]–[82]. The symetrized KLD is a measure of difference between the probability distributions $p(\mathbf{z}_q|\mathcal{H}_1)$ and $p(\mathbf{z}_q|\mathcal{H}_0)$ and is given by

$$\text{KLD}_q(\mathcal{H}_0, \mathcal{H}_1) = \frac{||\mathbb{E}_{\mathbf{z}_q|\mathcal{H}_1}(\mathbf{z}_q) - \mathbb{E}_{\mathbf{z}_q|\mathcal{H}_0}(\mathbf{z}_q)||^2}{\sigma_q^2} \tag{23}$$

for the considered binary hypothesis test. The above expression is derived in Appendix D.

According to the binary hypothesis detection test described by eq. (22) and the fact that $\mathbb{E}_{\mathbf{z}_q|\mathcal{H}_1}(\mathbf{z}_q) = \mathbf{D}_q \mathbf{s}_q$ and $\mathbb{E}_{\mathbf{z}_q|\mathcal{H}_0}(\mathbf{z}_q) = \mathbf{0}$, the symetrized KLD is given by the following simple expression $\text{KLD}_q(\mathcal{H}_0, \mathcal{H}_1) = \frac{||\mathbf{D}_q \mathbf{s}_q||^2}{\sigma_q^2}$. Inspecting the above expression, we can note that the $\text{KLD}_q(\mathcal{H}_0, \mathcal{H}_1)$ is in fact the output SNR_q[1]. Therefore, it makes sense to derive the optimal beamforming vector, denoted by $\mathbf{h}_q^{\text{opt.}}$, by maximizing the output SNR_q. More precisely, the following constrainted and convex optimization problem is considered:

$$\max_{\mathbf{h}_q} \ \mathbf{h}_q^T \mathbf{\Sigma}_q \mathbf{h}_q \ \text{ s.t. } ||\mathbf{h}_q||^2 = 1$$

with $\mathbf{\Sigma}_q = \mathbf{Q}_q^T \mathbf{D}_q^T \mathbf{D}_q \mathbf{Q}_q$ where $\mathbf{Q}_q$ is defined by eq. (11). The solution of the above problem is given by the eigenvector, $\mathbf{h}_q^{\text{opt}} = \mathcal{P}_{\max}(\mathbf{\Sigma}_q)$, associated with $\lambda_{\max}(\mathbf{\Sigma}_q)$. Note that $\mathbf{\Sigma}_q$ is a positive semi-definite rank-M_q matrix admitting the

[1]The $\text{KLD}_q(\mathcal{H}_0, \mathcal{H}_1)$ can be interpreted in the framework of the Generalized Likelihood Ratio Test (GLRT) [79]. Indeed, the GLRT, denoted by $T(\mathbf{z}_q)$, for unknown $\mathbf{s}_q$ decides $\mathcal{H}_1$ if $T(\mathbf{z}_q) > \gamma_q$ with a probability of false alarm given by $P_{\text{FA}} = Q_{\chi^2_{M_q}}(\gamma_q)$ and a probability of detection of $P_{\text{D}} = Q_{\chi^2_{M_q}(\text{SNR}_q)}(\gamma_q)$. Maximizing the non-centrality parameter SNR_q maximizes the hypothesis discrimination.

TABLE I
COST IN FLOPS FOR THE BOSE ASSOCIATED WITH THE OMP

Step	Operation	Cost for Q dimensions	Cost for $(M,\dots,M)$-block sparse cubic tensor
1st	Unfoldings	0	0
2nd	Beamforming	$\sum_{q=1}^{Q} O(\tilde{I}_q^2)$	$Q \cdot O(I^{2(Q-1)})$
3rd	Supports estim.	$\sum_{q=1}^{Q} O(M_q I_q K_q)$	$Q \cdot O(MIK)$
4th	LS estim.	$\sum_{q=1}^{Q} O(K_q M_q^2)$	$Q \cdot O(KM^2)$
Global cost	Proposed scheme	$\sum_{q=1}^{Q} O(M_q I_q (K_q + M_q) + \tilde{I}_q^2)$	$Q \cdot O(MI(K+M) + I^{2(Q-1)})$

Algorithm 1: Beamformed mOde based Sparse Estimator (BOSE).

Require: $\boldsymbol{\mathcal{Z}}$, $\{\mathbf{D}_1,\dots,\mathbf{D}_Q\}$, $\{\bar{M}_1,\dots,\bar{M}_Q\}$, any sparsity-based estimator $\mathcal{A}(\cdot)$ (*e.g.* OMP, CoSaMP, BP, Lasso, IHT,...)

for $1 \le q \le Q$ **do**
  $\mathbf{Z}^{(q)} = \mathrm{unfold}_q(\boldsymbol{\mathcal{Z}})$ ▷ q-mode unfolding
  $\hat{\mathbf{R}}_{\mathbf{Z}^{(q)}} = \frac{1}{I_q} \mathbf{Z}^{(q)^T} \mathbf{Z}^{(q)}$ ▷ Estimated SCM
  $\mathbf{h}_q^{\mathrm{opt.}} = \mathcal{P}_{\max}(\hat{\mathbf{R}}_{\mathbf{Z}^{(q)}})$ ▷ Optimal based beamforming
  $\mathbf{z}_q = \mathbf{Z}^{(q)} \mathbf{h}_q^{\mathrm{opt.}}$ ▷ Beamformed vector
  $\mathcal{A}(\mathbf{z}_q, \mathbf{D}_q, \bar{M}_q) \to \hat{\mathcal{F}}_q$ ▷ Estimation of the q-th support set
end for
Ensure: $\{\hat{\mathcal{F}}_1 \dots, \hat{\mathcal{F}}_Q\}$ ▷ Return of the Q estimated supports
$\hat{\boldsymbol{\mathcal{X}}}_{\hat{\mathcal{F}}} = \boldsymbol{\mathcal{Z}} \times_1 \mathbf{D}^{\dagger}_{\hat{\mathcal{F}}_1} \times_2 \dots \times_Q \mathbf{D}^{\dagger}_{\hat{\mathcal{F}}_Q}$ ▷ Least squares based estimation
Ensure: $\hat{\boldsymbol{\mathcal{X}}}_{\hat{\mathcal{F}}}$

following sorted eigendecomposition $\boldsymbol{\Sigma}_q = \mathbf{U}_q \mathbf{O}_q \mathbf{U}_q^T$ where $[\mathbf{O}_q]_{kk} = \lambda_k(\boldsymbol{\Sigma}_q)$ with $\lambda_k(\boldsymbol{\Sigma}_q) \ge \lambda_{k+1}(\boldsymbol{\Sigma}_q)$ and $\mathbf{U}_q$ is an orthonormal matrix. Unfortunately, the matrix $\mathbf{D}_q$, *i.e.*, the eigenspectrum of $\boldsymbol{\Sigma}_q$, is not directly available. But observe that the covariance matrix of $\mathbf{Z}^{(q)}$ admits the following eigendecomposition $\mathbf{R}_{\mathbf{Z}^{(q)}} = \mathbf{U}_q(\mathbf{O}_q^2 + \sigma^2\mathbf{I})\mathbf{U}_q^T$. Note that $\mathbf{R}_{\mathbf{Z}^{(q)}}$ and $\boldsymbol{\Sigma}_q$ share the same dominant right eigenspace denoted by $\mathcal{R}(\mathbf{U}_q)$. Now, it is straightforward to obtain $\lambda_{\max}(\boldsymbol{\Sigma}_q) = \max_k \sqrt{\lambda_k(\mathbf{R}_{\mathbf{Z}^{(q)}}) - \sigma^2}$. It is obvious that the above criterion is maximized for $\lambda_{\max}(\mathbf{R}_{\mathbf{Z}^{(q)}})$. Thus, picking the eigenvector associated with the largest eigenvalue of the matrix $\mathbf{R}_{\mathbf{Z}^{(q)}}$ leads to the selection of the desired optimal beamforming vector $\mathbf{h}_q^{\mathrm{opt}}$. In practice, we have to resort to the eigendecomposition of the sample covariance matrix (SCM) $\hat{\mathbf{R}}_{\mathbf{Z}^{(q)}} = \frac{1}{I_q}\mathbf{Z}^{(q)^T}\mathbf{Z}^{(q)}$ such that $\mathbf{h}_q^{\mathrm{opt.}} = \mathcal{P}_{\max}(\hat{\mathbf{R}}_{\mathbf{Z}^{(q)}})$. We can now formulate the BOSE scheme in Algorithm 1.

B. Evaluation of the Dominating Computational Cost in Flops

The OMP is known to have a relatively low computational cost as compared to other sparsity-based estimators. The dominating computational cost of the OMP on the $(\prod_{q=1}^{Q} I_q) \times (\prod_{q=1}^{Q} K_q)$ dictionary is $\prod_{q=1}^{Q} O(M_q I_q K_q)$ flops (FLoating-point OPerationS) [83]. In Table I, we summarize the cost of each step of the proposed method. The proposed method involves as an initial step, the unfolding operations which is essentially a reorganization of the data. The second step is the computation of the eigenvector associated to the largest eigenvalues of a SCM of size $\tilde{I}_q \times \tilde{I}_q$. Using, for instance, the orthogonal iteration method [84], this cost is evaluated as $O(\tilde{I}_q^2)$ flops. The third step is the OMP algorithm applied to each mode. This cost can be evaluated as $\sum_{q=1}^{Q} O(M_q I_q K_q)$ for the Q modes. Finally, the last step is the solution of Q ordinary LS problems. This final cost can be evaluated as $\sum_{q=1}^{Q} O(K_q M_q^2)$. So, we can conclude that the dominating cost of the proposed method is given by $\sum_{q=1}^{Q} O(M_q I_q (K_q + M_q))$. As by assumption, $K_q \gg M_q$, then the dominating cost is given by the third step.

V. APPLICATIONS TO TYPICAL OPERATIONAL PROBLEMS

To illustrate the proposed contributions, the two following important and challenging signal processing based applications are considered:

1) CS for non-bandlimited multidimensional signals,
2) CS-based channel estimation for MIMO wireless communication.

A. CS for Non-Bandlimited Multidimensional Signals

1) The 1D-Case: A typical real-valued time-continuous non-bandlimited signal is described by [56], [57], [85], [86]:

$$r(t) = \sum_{m=1}^{M} x_m \delta(t - \tau_m). \tag{24}$$

The vectors $\mathbf{x} = [x_1 \dots x_M]^T$ and $\boldsymbol{\tau} = [\tau_1 \dots \tau_M]^T$ denote the unknown amplitudes and the time-delays. Let $g(t)$ be a time-continuous sampling kernel. The regular sampling at rate $1/T$ of the signal $s(t)$ is given by

$$r_k = \int g(t - (k-1)T) r(t) dt = \sum_{m=1}^{M} x_m g(\tau_m - (k-1)T)$$

for $1 \le k \le K$. Define the $K \times K$ basis matrix $\boldsymbol{\Phi}$ by $[\boldsymbol{\Phi}]_{k,k'} = g((k' - k + 1)T)$ and the M-sparse vector $\bar{\mathbf{x}}$ of length K on support $\mathcal{F} = \{\tau_1, \ldots, \tau_M\}$. Then we have $\mathbf{z} = [z_1 \ldots z_K]^T = \mathbf{r} + \mathbf{w}$ where $\mathbf{w}$ is the noise in the digital domain, $\mathbf{r} = \mathbf{\Phi}\mathbf{x}^{\text{spar}} = \mathbf{\Phi}_{\mathcal{F}}\mathbf{x}_{\mathcal{S}}$ in which $[\mathbf{x}^{\text{spar}}]_{\mathcal{S}} = \mathbf{x}_{\mathcal{S}}$ and $[\mathbf{\Phi}_{\mathcal{F}}]_{k,m} = g(\tau_m - (k-1)T)$. Therefore, the sampled signal according to eq. (24) is sparse in the time domain.

2) Multidimensional QD-Sampling Extension:

a) Model definition: A QD-sampling is defined according to [58]:

$$r(t_1, \ldots, t_Q) = \sum_{m_1=1}^{M_1} \cdots \sum_{m_Q=1}^{M_Q} x_{m_1,\ldots,m_Q} \times \delta\Big(t_1 - \tau_{m_1}^{(1)}, \ldots, t_Q - \tau_{m_Q}^{(Q)}\Big). \quad (25)$$

Define the Q support sets as $\mathcal{F}_q = \{\tau_1^{(q)}, \ldots, \tau_{M_q}^{(q)}\}$ and let $\{T_1, \ldots, T_Q\}$ be Q sampling periods. The sampling coefficients arranged in a $K_1 \times \ldots \times K_Q$ tensor are given by eq. (26)

$$\begin{aligned}[\boldsymbol{\mathcal{R}}]_{k_1,\ldots,k_Q} &= \int \cdots \int g\Big(t_1 - (k_1-1)T_1, \ldots, t_Q - (k_Q-1)T_Q\Big) r(t_1, \ldots, t_Q) dt_1 \ldots dt_Q \\ &= \sum_{m_1=1}^{M_1} \cdots \sum_{m_Q=1}^{M_Q} x_{m_1,\ldots,m_Q} \prod_{q=1}^{Q} g(\tau_{m_q}^{(q)} - (k_q-1)T_q)\end{aligned} \quad (26)$$

for a separable sampling kernel. We formulate the following important remark.

Remark 5.1: $\boldsymbol{\mathcal{X}}^{\text{spar}}$ is a $(M_1, \ldots, M_Q)$-block sparse $K_1 \times \ldots \times K_Q$ tensor since $[\boldsymbol{\mathcal{X}}^{\text{spar}}]_{\mathcal{F}_1,\ldots,\mathcal{F}_Q} = \boldsymbol{\mathcal{X}}_{\mathcal{F}}$ where $\boldsymbol{\mathcal{X}}_{\mathcal{F}}$ contains the terms $x_{m_1,\ldots,m_Q}$ and is of size $M_1 \times \ldots \times M_Q$.

Result 5.2: The noisy $I_1 \times \ldots \times I_Q$ compressed measurement tensor for the QD-sampling of eq. (25) is given by eq. (7), where $[\mathbf{\Phi}_q]_{k_q,k'_q} = g((k'_q - k_q + 1)T_q)$ is the q-th $K_q \times K_q$ basis matrix. Thus, Result eq. (3.2) provides the CRB for the compressive QD-sampling with a separable sampling kernel.

b) Numerical illustrations: In this section, we assume a separable QD-Gaussian sampling kernel with $Q = 2$, defined by $g(t_1, \ldots, t_Q; \sigma_g^2) = \prod_{q=1}^{Q} g(t_q; \sigma_g^2)$ where $g(t_q; \sigma_g^2) = e^{-\frac{1}{2\sigma_g^2}(\frac{t_q}{T})^2}$ is a 1D-Gaussian kernel. The standard deviation σ_g determines the width of the kernel. The MSE curves are obtained by computing the trimmed/truncated mean over 500 Monte-Carlo trials of the squared error defined by $\text{SE}(trial) = ||\hat{\mathbf{x}}_{\hat{\mathcal{S}}}(trial) - \mathbf{x}_{\mathcal{S}}||^2$. In the trimmed/truncated mean, an equal amount of minimal and maximal squared errors are discarded. The number of rejected squared errors is usually given as a percentage of the total number of Monte-Carlo trials. In our context, this percentage is fixed to less than 1%. For a small percentage, the trimmed/truncated mean is well-known to be less sensitive to outliers but leaves the global tendency unchanged [87], [88]. The tensor $\boldsymbol{\mathcal{X}}$ is generated as a single realization of a multidimensional Gaussian distribution. So, $\boldsymbol{\mathcal{X}}^{\text{spar}} = \boldsymbol{\mathcal{X}} \bullet \boldsymbol{\mathcal{Q}}$ is random because of the random tensor $\boldsymbol{\mathcal{Q}}$. As a consequence, the SNR in dB is defined by averaging of $10\log_{10}\frac{||\boldsymbol{\mathcal{X}}^{\text{spar}}||^2}{K_1 K_2 \sigma^2}$ over the cardinalities.

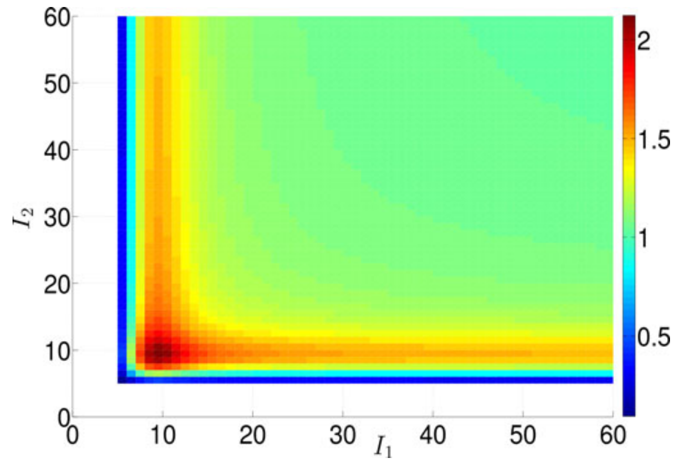

Fig. 3. Ratio $\frac{\mathcal{C}}{\mathcal{C}^\infty}$ *vs.* I_1 and I_2 where $I_1 \in [\bar{M}_1 + 1, 60]$, $I_2 \in [\bar{M}_2 + 1, 60]$, $K_1 = K_2 = 100$, $\bar{M}_1 = \bar{M}_2 = 4$.

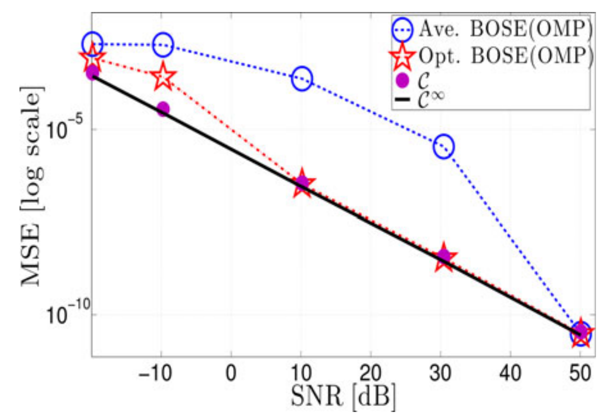

Fig. 4. MSEs *vs.* SNR with $I_1 = I_2 = 50$, $K_1 = 90$ and $K_2 = 100$ and $\bar{M}_1 = 2$ and $\bar{M}_2 = 3$.

In Fig. 3, the ratio of the lower bounds defined by $\frac{\mathcal{C}}{\mathcal{C}^\infty}$ and given by Results 3.2 and 3.4 is drawn as a function of the dimensions I_1 and I_2. We can observe that for growing I_1 and I_2, this ratio tends to one which illustrates the significance of the approximated closed-form expression $\mathcal{C}^\infty$.

In Fig. 4, the lower bound $\mathcal{C}$ is computed by the averaging of 500 Monte-Carlo trials of the expression given by eq. (27) and $\mathcal{C}^\infty$ is given in Result 3.4. We can note the good agreement of the numerically computed bound and the closed-form expression given in Result 3.4. In addition, the BOSE is associated with the OMP where the acronyms "Opt." or "Ave." indicate that the BOSE is derived with the max-SNR beamforming filter or with "column averaging," *i.e.*, $\mathbf{h}_q = [1/\sqrt{\tilde{I}_q} \ldots 1/\sqrt{\tilde{I}_q}]^T$ [76], [77]. We can see that the Opt. BOSE drastically outperforms the Ave. BOSE. In particular, the Opt. BOSE reaches the lower bounds at a much lower SNR than the Ave. BOSE.

In Fig. 5, the Opt. BOSE(OMP) is compared to the OMP estimator applied on the vectorized measurement tensor and on the Kronecker structured dictionary $\mathbf{D}_2 \otimes \mathbf{D}_1$ as defined in eq. (12). The popular Modified FOCal Underdetermined System Solver (MFOCUSS) estimator [36] is also tested. More precisely, we have $q - \text{MFOCUSS}(\mathbf{D}_q, \mathbf{Z}^{(q)}) \rightarrow (\hat{\mathbf{Q}}_q, \hat{\mathcal{F}}_q)$. Thanks to

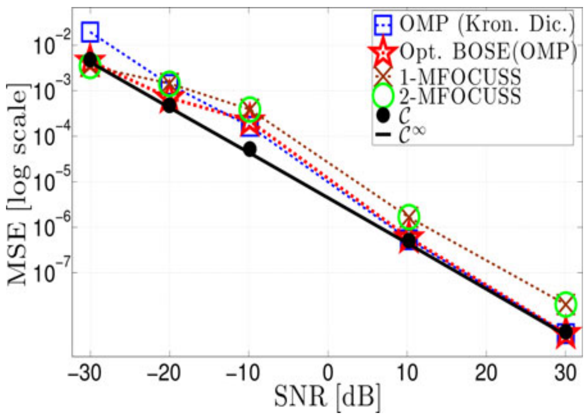

Fig. 5. MSEs *vs.* SNR with $I_1 = I_2 = 50$, $K_1 = K_2 = 100$, $\bar{M}_1 = \bar{M}_2 = 3$.

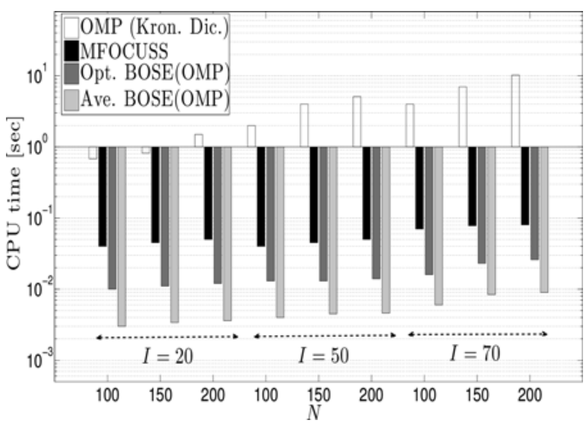

Fig. 6. CPU times in seconds and in log scale evaluated with MatLab.

eq. (10) and eq. (11), we have

$$\hat{\mathbf{X}}^{(q)}_{\hat{\mathcal{F}}} = \hat{\mathbf{Q}}_{\hat{\mathcal{F}}_q} \left(\mathbf{D}^T_{\hat{\mathcal{F}}_Q} \otimes \ldots \otimes \mathbf{D}^T_{\hat{\mathcal{F}}_{q+1}} \otimes \mathbf{D}^T_{\hat{\mathcal{F}}_{q-1}} \ldots \otimes \mathbf{D}^T_{\hat{\mathcal{F}}_1} \right)^\dagger$$

where $\hat{\mathbf{Q}}_{\hat{\mathcal{F}}_q}$ contains the M_q rows of $\hat{\mathbf{Q}}_q$ in the set $\hat{\mathcal{F}}_q$. We can see that this tensor-based extension of the MFOCUSS estimator exhibits a higher MSE than the BOSE. In addition, the Opt. BOSE(OMP) scheme globally outperforms the OMP with a Kronecker structured dictionary for a much lower computational cost.

The CPU times (in seconds) evaluated with MatLab for the tested algorithms are presented in Fig. 6 for several values of $I = I_1 = I_2$ and $N = N_1 = N_2$. We can see that the OMP based on a Kronecker dictionary has the highest computational cost, in particular for large I and/or N. Conversely, the cheaper methods is the BOSE(OMP) based on column averaging. Finally, BOSE(OMP) based on optimal beamforming shows a lower cost than the MFOCUSS algorithm and is approximatively equivalent to the BOSE(OMP) based on column averaging.

B. *Channel Estimation for MIMO Wireless Communication*

In the context of the MIMO wireless communication [89], we are interested in multi-path channel estimation in a pilot-assisted context. As illustrated in Fig. 7, pilot symbol sequences are sent

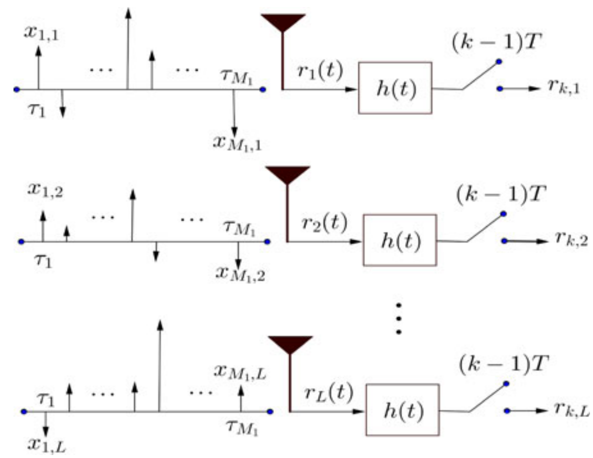

Fig. 7. Multipath channels in a MIMO wireless communication context.

by the transmitter to the multi-antenna receiver. The receiver in each channel performs a uniform sampling at rate $1/T$ thanks to a lowpass sampling kernel $h(t)$. For an *a priori* known time-delay sets $\{\tau_1, \ldots, \tau_{M_1}\}$, the aim is to estimate the time-varying coefficient $x_{m,\ell}$ in each channel resulting from different fading and shadowing effects.

1) Multichannel Compressive Sampling With a Common Support: According to [39], the considered application is well described by multichannel sampling according to a common support. More precisely, a multichannel time-continuous non-bandlimited signal is defined by

$$r_\ell(t) = \sum_{m=1}^{M_1} x_{m,\ell}\delta(t - \tau_m),$$

where $\ell \in \mathcal{L} = \{1, \ldots, L\}$ and L is the number of channels. The k-th sampled coefficients at rate $1/T$ in the ℓ-th channel is given by

$$\begin{aligned} r_{k,\ell} &= \int h(t - (k-1)T) r_\ell(t) dt \\ &= \sum_{m=1}^{M_1} x_{m,\ell} h(\tau_m - (k-1)T). \end{aligned}$$

This multichannel model exhibits a common support $\mathcal{F}_1 = \{\tau_1, \ldots, \tau_{M_1}\}$ in each channel (see Fig. 7).

Remark 5.3: The $K_1 \times K_1$ basis matrix $\mathbf{\Phi}_1$ is defined by the sampling of the kernel $h(t)$. Then, the following expression holds

$$\mathbf{R} = \mathbf{X}^{\text{spar}} \times_1 \mathbf{\Phi}_1 \times_2 \mathbf{I}_L = \mathbf{\Phi}_{\mathcal{F}_1} \mathbf{X}_{\mathcal{F}_1}$$

where $\mathbf{R}$ is $K_1 \times L$ and $\mathbf{X}^{\text{spar}}$ is (M_1, L)-block sparse since $[\mathbf{X}^{\text{spar}}]_{\mathcal{F}_1, \mathcal{L}} = \mathbf{X}_{\mathcal{F}_1}$ where $\mathbf{X}_{\mathcal{F}_1}$ is $M_1 \times L$.

Now, consider the CS framework and let $\mathbf{\Psi}_1$ be a measurement matrix of size $I_1 \times K_1$ with $I_1 < K_1$. We have

$$\mathbf{Y} = \mathbf{\Psi}_1 \mathbf{R} = \mathbf{\Psi}_1 \mathbf{\Phi}_1 \mathbf{X}^{\text{spar}} \mathbf{I}_L = \mathbf{D}_1 \mathbf{X}^{\text{spar}}.$$

Now, we are ready to describe the noisy $(I_1 L) \times 1$ measurement vector according to

$$\mathbf{z} = (\mathbf{I}_L \otimes \mathbf{D}_1)\mathbf{x}^{\text{spar}} + \mathbf{w}$$

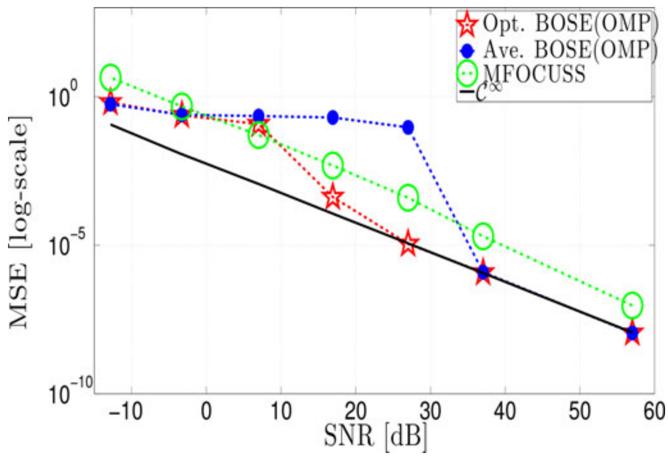

Fig. 8. MSEs *vs.* SNR with $L = 3$ channels, $I_1 = 30$, $K_1 = 70$, $\bar{M}_1 = 3$.

where $(\mathbf{I}_L \otimes \mathbf{D}_1)\mathbf{x}^{\text{spar}} = (\mathbf{I}_L \otimes \mathbf{D}_{\mathcal{F}_1})\mathbf{x}_{\mathcal{S}}$. In the multichannel context, we assume that the noise in each channel can have a different variance, denoted by σ_ℓ^2. More precisely, we have $\mathbb{E}(w_{k,\ell} w_{k',\ell'}) = \sigma_\ell^2 \delta_{\ell-\ell'} \delta_{k-k'}$ where $\delta_.$ is the Kronecker delta symbol and $w_{k,\ell}$ is the noise sample in the ℓ-th channel for the k-th sample.

Result 5.4: Under the above assumptions and the model of Section III-A and as

$$\mathbf{z}|\mathbf{D}_{\mathcal{F}_1}, \mathbf{x}_{\mathcal{S}} \sim \mathcal{N}\left((\mathbf{I}_L \times \mathbf{D}_{\mathcal{F}_1})\mathbf{x}_{\mathcal{S}}, \sigma^2 \mathbf{I}_L\right),$$

where $\sigma^2 \mathbf{I}_L = \text{diag}\{\sigma_1^2, \dots, \sigma_L^2\} \otimes \mathbf{I}_{I_1}$, Result 3.2 holds with $Q = 1$ and $\sigma^2 = \sum_{\ell=1}^{L} \sigma_\ell^2$.

2) Numerical Illustrations: In this section, we assume a sinus cardinal sampling kernel defined as $h(t) = \frac{1}{T}\text{sinc}\left(\frac{t}{T}\right)$ which is an ideal lowpass filter with frequency support $[-\pi/T, \pi/T]$ [23]. According to Fig. 8, we can see the efficiency of the BOSE(OMP) based on optimal beamforming.

VI. Conclusion

Compressive Sampling (CS) of sparse signals allows to overcome the limits of Shannon's sampling theorem. In the era of big data, tensors provide a natural representation for such massive multidimensional data and the concept of block-sparse core tensors is the multidimensional generalization of the concept of block sparsity for vectors. In this work, it is assumed that each unknown entry of a block-sparse core tensor is Bernoulli-distributed, meaning that each entry has a given probability to be non-zero. This statistical prior leads to a set of random supports of truncated Binomial-distributed cardinalities. In this context a Cramér-Rao lower Bound (CRB) on the Mean Squared Error (MSE) of the estimated non-zero entries of the core tensor is derived in a compact form for Gaussian measurement matrices. Based on the limit form of the Poisson distribution, an approximated CRB expression is provided for large dictionaries and a highly block-sparse core tensor. The second part of this work is dedicated to the proposition of the Beamformed mOde based Sparse Estimator (BOSE) which exploits the property that the mode unfolding matrices of a block-sparse tensor follow the Multiple-Measurement Vectors (MMV) model with joint sparsity patterns. More precisely, the initial step of the BOSE is to "map" the MMV model onto the Single MV model thanks to a beamforming filter. Next, any standard sparsity-based estimator adapted to the Single MV model can be associated to the BOSE. The estimation performance of the BOSE and its statistical efficiency are illustrated by means of numerical simulations in the context of CS of non-bandlimited multidimensional signals and for MIMO wireless channel communication. As a straightforward perspective, the BOSE seems well adapted to the sparse source localization in array processing.

Appendix

A. Proof of Result 3.2

It is well-known that the CRB [40], [87] is a lower bound of the MSE for an unbiased estimator according to

$$\text{MSE} \geq \text{CRB} = \frac{1}{\bar{I}}\text{Tr}\left[\left(\text{Var}\left(\frac{\partial \log p(\mathbf{z}, \mathbf{D}_{\mathcal{F}}|\mathbf{x}_{\mathcal{S}})}{\partial \mathbf{x}_{\mathcal{S}}}\right)\right)^{-1}\right]$$

where the joint log-pdf of class C^1 is given by $\log p(\mathbf{z}, \mathbf{D}_{\mathcal{F}}|\mathbf{x}_{\mathcal{S}}) = \log p(\mathbf{z}|\mathbf{x}_{\mathcal{S}}, \mathbf{D}_{\mathcal{F}}) + \log p(\mathbf{D}_{\mathcal{F}})$. As $\mathbf{D}_{\mathcal{F}}$ is not a function of $\mathbf{x}_{\mathcal{S}}$, the score function can be simplified according to $\frac{\partial \log p(\mathbf{z}, \mathbf{D}_{\mathcal{F}}|\mathbf{x}_{\mathcal{S}})}{\partial \mathbf{x}_{\mathcal{S}}} = \frac{\partial \log p(\mathbf{z}|\mathbf{x}_{\mathcal{S}}, \mathbf{D}_{\mathcal{F}})}{\partial \mathbf{x}_{\mathcal{S}}}$. Consequently the averaged MSE is lower bounded by the averaged CRB given by

$$\begin{aligned}\text{MSE}_{av.} \geq \mathcal{C} = {} & \mathbb{E}_{\mathcal{M}}\mathbb{E}_{\mathbf{D}_{\mathcal{F}}|\mathcal{M}}\text{CRB} \\ = {} & \frac{1}{\bar{I}}\mathbb{E}_{\mathcal{M}}\mathbb{E}_{\mathbf{D}_{\mathcal{F}}|\mathcal{M}} \\ & \times \text{Tr}\left[\left(\text{Var}\left(\frac{\partial \log p(\mathbf{z}|\mathbf{D}_{\mathcal{F}}, \mathbf{x}_{\mathcal{S}})}{\partial \mathbf{x}_{\mathcal{S}}}\right)\right)^{-1}\right].\end{aligned}$$

The statistics of the measurement vector are $\mathbf{z}|\mathbf{D}_{\mathcal{F}}, \mathbf{x}_{\mathcal{S}} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$, where $\boldsymbol{\mu} = \mathbb{E}(\mathbf{z}|\mathbf{D}_{\mathcal{F}}, \mathbf{x}_{\mathcal{S}}) = \mathbf{D}_{\mathcal{F}}\mathbf{x}_{\mathcal{S}}$. So, using the Slepian-Bang formula [41], we obtain

$$\text{Tr}\left[\left(\text{Var}\left(\frac{\partial \log p(\mathbf{z}|\mathbf{D}_{\mathcal{F}}, \mathbf{x}_{\mathcal{S}})}{\partial \mathbf{x}_{\mathcal{S}}}\right)\right)^{-1}\right] = \text{Tr}\left[\mathbf{F}^{-1}\right]$$

where the Fisher Information Matrix (FIM) is given by

$$\begin{aligned}\mathbf{F}_{\mathcal{S}} &= \frac{1}{\sigma^2}\left(\frac{\partial \boldsymbol{\mu}}{\partial \mathbf{x}_{\mathcal{S}}}\right)^T \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{x}_{\mathcal{S}}} = \frac{1}{\sigma^2}\mathbf{D}_{\mathcal{F}}^T\mathbf{D}_{\mathcal{F}} \\ &= \frac{1}{\sigma^2}\mathbf{D}_{\mathcal{F}_1}^T\mathbf{D}_{\mathcal{F}_1} \otimes \ldots \otimes \mathbf{D}_{\mathcal{F}_Q}^T\mathbf{D}_{\mathcal{F}_Q}.\end{aligned}$$

Using the property of the Kronecker product with respect to the matrix inversion (see Lemma 2.3), the inverse of the FIM is easily obtained according to

$$\mathbf{F}_{\mathcal{S}}^{-1} = \frac{1}{\sigma^2}\left(\mathbf{D}_{\mathcal{F}_1}^T\mathbf{D}_{\mathcal{F}_1}\right)^{-1} \otimes \ldots \otimes \left(\mathbf{D}_{\mathcal{F}_Q}^T\mathbf{D}_{\mathcal{F}_Q}\right)^{-1}.$$

Thus, using the property of the Kronecker product with respect to the trace operator (see Lemma 2.3), we obtain

$$\begin{aligned}\mathcal{C} =& \frac{\sigma^2}{\bar{I}}\mathbb{E}_{\mathcal{M}}\mathbb{E}_{\mathbf{D}_{\mathcal{F}}|\mathcal{M}}\mathrm{Tr}\left[\mathbf{F}_{\mathcal{S}}^{-1}\right] \\ =& \sigma^2\mathbb{E}_{\mathcal{M}}\prod_{q=1}^{Q}\frac{1}{I_q}\mathbb{E}_{\mathbf{D}_{\mathcal{F}_q}|\mathcal{M}}\mathrm{Tr}\left[(\mathbf{D}_{\mathcal{F}_q}^T\mathbf{D}_{\mathcal{F}_q})^{-1}\right] \\ =& \sigma^2\mathbb{E}_{\mathcal{M}}\prod_{q=1}^{Q}\mathbb{E}_{\mathbf{D}_{\mathcal{F}_q}|\mathcal{M}}\mathrm{Tr}\left[(\bar{\mathbf{D}}_{\mathcal{F}_q}^T\bar{\mathbf{D}}_{\mathcal{F}_q})^{-1}\right], \end{aligned} \tag{27}$$

where the i.i.d. entries of the matrix $\bar{\mathbf{D}}_{\mathcal{F}_q} = \sqrt{I_q}\mathbf{D}_{\mathcal{F}_q}$ follows the distribution $\mathcal{N}(0,1)$. Using [90], [91], we get

$$\mathbb{E}_{\mathbf{D}_{\mathcal{F}_q}}\mathrm{Tr}\left[(\bar{\mathbf{D}}_{\mathcal{F}_q}^T\bar{\mathbf{D}}_{\mathcal{F}_q})^{-1}\right] = \frac{M_q}{I_q - M_q}.$$

Using the independence assumption of the random variable M_q and taking into account the model identifiability constraint, the CRB is given by

$$\begin{aligned}\mathcal{C} =& \sigma^2\prod_{q=1}^{Q}\mathbb{E}_{\mathcal{M}_q}\left(\frac{M_q}{I_q - M_q}\right) \\ =& \sigma^2\prod_{q=1}^{Q}\sum_{m_q=1}^{K_q}\frac{m_q}{I_q - m_q}\mathrm{Pr}\left(M_q = m_q \mid M_q \in \mathcal{V}_q\right) \\ =& \sigma^2\prod_{q=1}^{Q}\frac{1}{P_q^{\mathrm{id}}}\sum_{m_q=1}^{K_q}\frac{m_q}{I_q - m_q}\mathrm{Pr}(M_q = m_q)1_{\mathcal{V}_q}(m_q) \\ =& \sigma^2\prod_{q=1}^{Q}\frac{1}{P_q^{\mathrm{id}}}\sum_{m_q\in\mathcal{V}_q}\frac{m_q}{I_q - m_q}\mathrm{Pr}(M_q = m_q).\end{aligned}$$

B. Approximation of $\mathcal{P}_q^{\mathrm{id}}$

Recalling the definition of P_q^{id} given in eq. (16) and using the binomial formula, we have

$$\mathcal{P}_q^{\mathrm{id}} = \lim_{K_q\to\infty}P_q^{\mathrm{id}} = 1 - \lim_{K_q\to\infty}\mathrm{Pr}\big(M_q = 0\big) - \epsilon(I_q) \tag{28}$$

where $\epsilon(I_q)$ has been defined previously and

$$\mathrm{Pr}\big(M_q = 0\big) = (1 - P_q)^{K_q} = e^{K_q\ln(1-P_q)} = e^{K_q(-P_q + O(P_q^2))}$$

using the Taylor expansion of the natural logarithm. Recall that P_q has to be sufficiently small to satisfy $\lim_{K_q\to\infty}K_qP_q = \bar{M}_q < \infty$. Thus, it is realistic to discard the term in $O(P_q^2)$ in the Taylor expansion of function $\ln(1-P_q)$. By doing this, we obtain

$$\lim_{K_q\to\infty}\mathrm{Pr}\big(M_q = 0\big) \approx \lim_{K_q\to\infty}e^{-K_qP_q} = e^{-\bar{M}_q}.$$

Using the above approximation and eq. (28) lead to eq. (20).

C. Proof of Result 3.4

Lemma 6.1: For $X \sim$ Poisson (θ), we have

$$\mathbb{E}\left(\frac{X}{N - X}\right) \approx \frac{\theta}{(N-\theta-2)(1-e^{-(N-\theta-1)})}. \tag{29}$$

Proof:

1) According to [92], for $X \sim$ Poisson (θ) we have $\mathbb{E}\left(Xf(X)\right) = \theta\mathbb{E}\left(f(X+1)\right)$ for every bounded function $f(\cdot)$ on a given domain.
2) In addition, the inverse moment of $Y \sim$ Poisson (θ') can be approximated according to $\mathbb{E}\left(Y^{-1}\right) \approx \frac{1}{(\theta'-1)((1-e^{-\theta'})}$ [93].

So, consider the bounded function $f(X) = \frac{1}{N-X}$ on $\{0,\dots,N-1\}$. Then, we have $\mathbb{E}\left(\frac{X}{N-X}\right) = \theta\mathbb{E}\left(\frac{1}{N-X-1}\right)$. Now let $Y = N - X - 1$, then $\mathbb{E}\left(\frac{1}{N-X-1}\right) = \mathbb{E}\left(Y^{-1}\right)$ for $Y \sim$ Poisson (θ'). Thus $\mathbb{E}\left(\frac{X}{N-X}\right) = \theta\mathbb{E}\left(Y^{-1}\right)$ with $\theta' = N - \mathbb{E}X - 1 = N - \theta - 1$ give eq. (29). ■

The CRB for $K_1,\dots,K_Q \to \infty$ is defined as

$$\begin{aligned}\mathcal{C}^{\infty} =& \lim_{K_1,\dots,K_Q\to\infty}\mathcal{C} \\ =& \sigma^2\prod_{q=1}^{Q}\lim_{K_q\to\infty}\frac{1}{P_q^{\mathrm{id}}}\sum_{m_q\in\mathcal{V}_q}\frac{m_q}{I_q - m_q}\mathrm{Pr}(M_q = m_q) \\ =& \sigma^2\prod_{q=1}^{Q}\frac{1}{\mathcal{P}_q^{\mathrm{id}}}\sum_{m_q\in\mathcal{V}_q}\frac{m_q}{I_q - m_q}\lim_{K_q\to\infty}\mathrm{Pr}(M_q = m_q) \\ =& \sigma^2\prod_{q=1}^{Q}\frac{1}{\mathcal{P}_q^{\mathrm{id}}}\sum_{m_q\in\mathcal{V}_q}\frac{m_q}{I_q - m_q}\frac{\bar{M}_q^{m_q}}{m_q!}e^{-\bar{M}_q}\end{aligned}$$

using the limit form of the Poisson distribution given by eq. (19).

For sufficiently large $I_q \gg 1$, $\mathcal{P}_q^{\mathrm{id}} \approx 1 - e^{-\bar{M}_q}$ (see the second point of Remark 3.3), and using Lemma 6.1 for $M_q \overset{a}{\sim}$ Poisson $(\bar{M}_q)$, we get the following approximation:

$$\begin{aligned}\mathcal{C}^{\infty} \approx& \sigma^2\prod_{q=1}^{Q}\frac{1}{1-e^{-\bar{M}_q}}\mathbb{E}_{M_q}\left(\frac{M_q}{I_q - M_q}\right) \\ \approx& \sigma^2\prod_{q=1}^{Q}\frac{1}{(1-e^{-\bar{M}_q})(1-e^{-(I_q-\bar{M}_q-1)})}\frac{\bar{M}_q}{I_q-\bar{M}_q-2} \\ \approx& \sigma^2\prod_{q=1}^{Q}\frac{\bar{M}_q}{(1-e^{-\bar{M}_q})(1-e^{-I_q}e^{\bar{M}_q})(I_q-\bar{M}_q)}.\end{aligned}$$

D. Symetrized KLD for Real Gaussian pdf

Consider the following binary hypothesis test:

$$\begin{cases}\mathcal{H}_0: \ \mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}_0, \mathbf{R}_0) \\ \mathcal{H}_1: \ \mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}_1, \mathbf{R}_1)\end{cases}$$

where $\boldsymbol{\mu}_i = \mathbb{E}_{\mathbf{z}|\mathcal{H}_i}(\mathbf{z})$ and $\mathbf{R}_i$ are $L \times L$ non-singular covariance matrices under hypothesis $\mathcal{H}_i$. Therefore, the problem of interest is the discrimination of multidimensional real Gaussian processes. The two KLDs for real Gaussian pdfs [59] are given by

$$\begin{aligned}\mathrm{KLD}(\mathcal{H}_1|\mathcal{H}_0) =& \int p(\mathbf{z}|\mathcal{H}_1)\log\frac{p(\mathbf{z}|\mathcal{H}_1)}{p(\mathbf{z}|\mathcal{H}_0)}d\mathbf{z}\\ =& \frac{1}{2}\Bigg(\mathrm{Tr}[\mathbf{R}_0^{-1}\mathbf{R}_1] - L + (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^T\\ &\times \mathbf{R}_0^{-1}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1) + \ln\frac{\det(\mathbf{R}_0)}{\det(\mathbf{R}_1)}\Bigg)\\ \mathrm{KLD}(\mathcal{H}_0|\mathcal{H}_1) =& \int p(\mathbf{z}|\mathcal{H}_0)\log\frac{p(\mathbf{z}|\mathcal{H}_0)}{p(\mathbf{z}|\mathcal{H}_1)}d\mathbf{z}\\ =& \frac{1}{2}\Bigg(\mathrm{Tr}[\mathbf{R}_1^{-1}\mathbf{R}_0] - L + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\\ &\times \mathbf{R}_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) + \ln\frac{\det(\mathbf{R}_1)}{\det(\mathbf{R}_0)}\Bigg).\end{aligned}$$

Generally $\mathrm{KLD}(\mathcal{H}_0|\mathcal{H}_1)$ and $\mathrm{KLD}(\mathcal{H}_1|\mathcal{H}_0)$ are not symmetric with respect to the hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$. As a consequence, the KLD cannot serve as a distance. Despite of this technical problem, the KLD is a central pseudo-metric intensively used in many real applications. This can be partially explained by the existing strong link between the KLD and the optimal Neyman-Pearson test [94]. To alleviate this problem, it is usual to consider the symetrized KLD [82] defined by

$$\begin{aligned}\mathrm{KLD}(\mathcal{H}_0,\mathcal{H}_1) =& \mathrm{KLD}(\mathcal{H}_1|\mathcal{H}_0) + \mathrm{KLD}(\mathcal{H}_0|\mathcal{H}_1)\\ =& \frac{1}{2}\Bigg(\mathrm{Tr}[\mathbf{R}_0^{-1}\mathbf{R}_1 + \mathbf{R}_1^{-1}\mathbf{R}_0] - 2L\\ &+ (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\mathbf{R}_1^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\\ &+ (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^T\mathbf{R}_0^{-1}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)\\ &+ \ln\frac{\det(\mathbf{R}_1)}{\det(\mathbf{R}_0)} + \ln\frac{\det(\mathbf{R}_0)}{\det(\mathbf{R}_1)}\Bigg)\\ =& \frac{1}{2}\mathrm{Tr}[\mathbf{R}_0^{-1}\mathbf{R}_1 + \mathbf{R}_1^{-1}\mathbf{R}_0] - L\\ &+ (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T\frac{\mathbf{R}_1^{-1} + \mathbf{R}_0^{-1}}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0).\end{aligned}$$

Now assume that $\mathbf{R}_0 = \mathbf{R}_1 = \sigma^2\mathbf{I}$, then the symetrized KLD simplifies to

$$\mathrm{KLD}(\mathcal{H}_0,\mathcal{H}_1) = \frac{||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0||^2}{\sigma^2}$$

which coincides with eq. (23).

References

[1] P. M. Kroonenberg, *Applied Multiway Data Analysis*, vol. 702. Hoboken, NJ, USA: Wiley, 2008.

[2] P. Comon, "Tensors: A brief introduction," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 44–53, May 2014.

[3] N. D. Sidiropoulos, "Low-rank decomposition of multi-way arrays: A signal processing perspective," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop*, 2004, pp. 52–58.

[4] A. Cichocki *et al.*, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.

[5] L. De Lathauwer, "A survey of tensor methods," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2009, pp. 2773–2776.

[6] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[7] N. Sidiropoulos, E. E. Papalexakis, and C. Faloutsos, "Parallel randomly compressed cubes: A scalable distributed architecture for big tensor decomposition," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 57–70, Sep. 2014.

[8] A. Cichocki, "Era of big data processing: A new approach via tensor networks and tensor decompositions," in *Proc. Int. Workshop Smart Info-Media Syst. Asia* (SISA-2013), 2013.

[9] L. Kuang, F. Hao, L. T. Yang, M. Lin, C. Luo, and G. Min, "A tensor-based approach for big data representation and dimensionality reduction," *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 3, pp. 280–291, Sep. 2014.

[10] R. Badeau and R. Boyer, "Fast multilinear singular value decomposition for structured tensors," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1008–1021, 2008.

[11] M. Boizard, R. Boyer, G. Favier, and P. Larzabal, "Fast multilinear singular value decomposition for higher-order Hankel tensors," in *Proc. IEEE 8th Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2014, pp. 437–440, doi: 10.1109/SAM.2014.6882436.

[12] S. Ragnarsson, "Structured tensor computations: Blocking, symmetries and kronecker factorizations," Ph.D. dissertation, Cornell Univ., Ithaca, NY, USA, 2012.

[13] W. Ding, L. Qi, and Y. Wei, "Fast Hankel tensor-vector product and its application to exponential data fitting," *Numer. Linear Algebra Appl.*, vol. 22, pp. 814–832, 2015.

[14] L. Qi, Q. Wang, and Y. Chen, "Three dimensional strongly symmetric circulant tensors," *Linear Algebra Appl.*, vol. 482, pp. 207–220, 2015.

[15] R. Boyer, R. Badeau, and G. Favier, "Fast orthogonal decomposition of volterra cubic kernels using oblique unfolding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2011, pp. 4080–4083.

[16] J. H. de Morais Goulart, M. Boizard, R. Boyer, G. Favier, and P. Comon, "Tensor CP decomposition with structured factor Matrices: Algorithms and performance," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 4, pp. 757–769, Jun. 2016.

[17] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3198–3213, Jul. 2008.

[18] D. Nion and N. D. Sidiropoulos, "Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5693–5705, Nov. 2010.

[19] R. Boyer, "Deterministic asymptotic Cramer–Rao bound for the multidimensional harmonic model," *Signal Process.*, vol. 88, no. 12, pp. 2869–2877, 2008.

[20] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[21] R. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[22] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[23] M. Unser, "Sampling-50 years after Shannon," in *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.

[24] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[25] M. Akçakaya and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 492–504, Jan. 2010.

[26] Y. Wang, G. Leus, and A. Pandharipande, "Direction estimation using compressive sampling array processing," in *Proc. IEEE 15th Workshop Statist. Signal Process.*, 2009, pp. 626–629.

[27] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proc. 5th Int. Conf. Inf. Process. Sensor Netw.*, 2006, pp. 134–142.

[28] V. Stankovic, L. Stankovic, and S. Cheng, "Compressive video sampling," in *Proc. 16th Eur. Signal Process. Conf.*, 2008, pp. 1–5.

[29] Y. Yu, A. P. Petropulu, and H. V. Poor, "Mimo radar using compressive sampling," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 146–163, Feb. 2010.

[30] S. Friedland, Q. Li, and D. Schonfeld, "Compressive sensing of sparse tensors," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4438–4447, Oct. 2014.

[31] N. D. Sidiropoulos and A. Kyrillidis, "Multi-way compressed sensing for sparse low-rank tensors," *IEEE Signal Process. Lett.*, vol. 19, no. 11, pp. 757–760, Nov. 2012.

[32] C. F. Caiafa and A. Cichocki, "Multidimensional compressed sensing and their applications," *Wiley Interdisciplinary Rev., Data Mining Knowl. Discovery*, vol. 3, no. 6, pp. 355–380, 2013.

[33] L.-H. Lim and P. Comon, "Multiarray signal processing: Tensor decomposition meets compressed sensing," *Comptes Rendus Mecanique*, vol. 338, no. 6, pp. 311–320, 2010.

[34] Y. C. Eldar, P. Kuppinger, and H. Bölcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[35] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, Aug. 2009.

[36] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[37] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2006.

[38] J. M. Kim, O. K. Lee, and J. C. Ye, "Compressive MUSIC: Revisiting the link between compressive sensing and array signal processing," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 278–301, Jan. 2012.

[39] A. Hormati and M. Vetterli, "Compressive sampling of multiple sparse signals having common support using finite rate of innovation principles," *IEEE Signal Process. Lett.*, vol. 18, no. 5, pp. 331–334, May 2011.

[40] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[41] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Prentice-Hall, 2005.

[42] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, "On the achievability of Cramer–Rao bound in noisy compressed sensing," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 518–526, Jan. 2012.

[43] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramér–Rao bound for noisy compressive sampling," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1233–1236, Mar. 2009.

[44] Z. Ben-Haim and Y. Eldar, "The Cramèr-Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3384–3389, Jun. 2010.

[45] R. Prasad and C. R. Murthy, "Cramér-Rao-type bounds for sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 622–632, Feb. 2013.

[46] R. Boyer, P. Larzabal, and B. H. Fleury, "Oracle performance estimation of Bernoulli-distributed sparse vectors," in *Proc. IEEE Statistical Signal Process. Workshop (SSP)*, Palma de Mallorca, Spain, 2016, pp. 1–4, doi: 10.1109/SSP.2016.7551780.

[47] S. Bernhardt, R. Boyer, S. Marcos, and P. Larzabal, "Compressed sensing with basis mismatch: Performance bounds and sparse-based estimator," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3483–3494, Jul. 2016.

[48] S. Sahnoun and P. Comon, "Joint source estimation and localization," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2485–2495, May 2015.

[49] X. Liu and N. D. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 2074–2086, Sep. 2001.

[50] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Conf. Rec. 27th Asilomar Conf.*, 1993, pp. 40–44.

[51] J. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[52] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[53] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[54] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc B Methodol.*, vol. 58, pp. 267–288, 1996.

[55] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.

[56] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, Jun. 2002.

[57] R. Tur, Y. C. Eldar, and Z. Friedman, "Innovation rate sampling of pulse streams with application to ultrasound imaging," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1827–1842, Apr. 2011.

[58] P. Shukla and P. L. Dragotti, "Sampling schemes for multidimensional signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3670–3686, Jul. 2007.

[59] N. L. Johnson, A. W. Kemp, and S. Kotz, *Univariate Discrete Distributions*, vol. 444. Hoboken, NJ, USA: Wiley, 2005.

[60] K. Lee, Y. Bresler, and M. Junge, "Subspace methods for joint sparse recovery," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3613–3641, Jun. 2012.

[61] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[62] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. New York, NY, USA: Springer-Verlag, 2013.

[63] E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 1, pp. 59–73, 2011.

[64] M. A. Davenport, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "A simple proof that random matrices are democratic," Tech. Rep., arXiv:0911.0736, 2009.

[65] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, vol. 1. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

[66] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.

[67] F. Roemer, G. Del Galdo, and M. Haardt, "Tensor-based algorithms for learning multidimensional separable dictionaries," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2014, pp. 3963–3967.

[68] C.-S. Lu and W.-J. Liang, "Fast compressive sensing of high-dimensional signals with tree-structure sparsity pattern," in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process.*, 2014, pp. 738–742.

[69] R. Rubinstein, M. Zibulevsky, and M. Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1553–1564, Mar. 2010.

[70] M. Protter, I. Yavneh, and M. Elad, "Closed-form MMSE estimation for signal denoising under sparse representation modeling over a unitary dictionary," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3471–3484, Jul. 2010.

[71] S. Som and L. C. Potter, "Sparsity pattern recovery in Bernoulli-Gaussian signal model," Tech. Rep., arXiv:1004.4044, 2010.

[72] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, Jan. 2010.

[73] M. Smith and R. Kohn, "Nonparametric regression using Bayesian variable selection," *J. Econometr.*, vol. 75, no. 2, pp. 317–343, 1996.

[74] P. Schniter, L. C. Potter, and J. Ziniel, "Fast Bayesian matching pursuit," in *Proc. IEEE Inf. Theory Appl. Workshop*, 2008, pp. 326–333.

[75] N. Dobigeon and J.-Y. Tourneret, "Bayesian orthogonal component analysis for sparse representation," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2675–2685, May 2010.

[76] D. Malioutov, "A sparse signal reconstruction perspective for source localization with sensor arrays," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., 2003. [Online]. Available: http://ssg.mit.edu/dmm/publications/

[77] N. Wagner, Y. C. Eldar, and Z. Friedman, "Compressed beamforming in ultrasound imaging," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4643–4657, Sep. 2012.

[78] J. Li and P. Stoica, *Robust Adaptive Beamforming*. New York, NY, USA: Wiley Online Library, 2006.

[79] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. 1. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.

[80] S. Sinanović and D. H. Johnson, "Toward a theory of information processing," *Signal Process.*, vol. 87, no. 6, pp. 1326–1344, 2007.

[81] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: WIley, 1991.

[82] P. Kumar and A. Johnson, "On a symmetric divergence measure and information inequalities," *J. Inequalities Pure Appl. Math.*, vol. 6, no. 3, 2005. Art. no. 65.

[83] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6202–6216, Dec. 2012.

[84] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 3. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 2012.

[85] S. Bernhardt, R. Boyer, S. Marcos, Y. Eldar, and P. Larzabal, "Cramer-Rao Bound for finite streams of an arbitrary number of pulses," in *Conf. Proc. EUSIPCO*, Lisbonne, Portugal, Sep. 2014. [Online]. Available: https://hal-supelec.archives-ouvertes.fr/hal-01005005

[86] S. Bernhardt, R. Boyer, S. Marcos, Y. C. Eldar, and P. Larzabal, "Sampling FRI signals with the SOS kernel: Bounds and optimal kernel," in *Proc. 23rd Eur. Signal Process. Conf.*, Aug. 2015, pp. 2172–2176.

[87] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, vol. 31. New York, NY, USA: Springer Science & Business Media, 1998.

[88] A. DasGupta, "The trimmed mean," in *Asymptotic Theory of Statistics and Probability* (ser. Springer Texts in Statistics). New York, NY, USA: Springer-Verlag, 2008, pp. 271–278.

[89] J. G. Proakis, M. Salehi, N. Zhou, and X. Li, *Communication Systems Engineering*, vol. 94. Englewood Cliffs, NJ, USA: Prentice-Hall, 1994.

[90] A. Lozano, A. M. Tulino, and S. Verdú, "Multiple-antenna capacity in the low-power regime," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2527–2544, Oct. 2003.

[91] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*, 1st ed. New York, NY, USA: Cambridge Univ. Press, 2011.

[92] P. Diaconis and S. Zabell, "Closed form summation for classical distributions: Variations on a theme of de Moivre," *Statist. Sci.*, vol. 6, pp. 284–302, 1991.

[93] M. Tiku, "A note on the negative moments of a truncated Poisson variate," *J. Amer. Statist. Assoc.*, vol. 59, no. 308, pp. 1220–1224, 1964.

[94] S. Eguchi and J. Copas, "Interpreting Kullback–Leibler divergence with the Neyman–Pearson lemma," *J. Multivar. Anal.*, vol. 97, no. 9, pp. 2034–2040, 2006.

Rémy Boyer received the M.Sc. and Ph.D. degrees from the Ecole Nationale Suprieure des Telecommunications, Paris, France, in 1999 and 2002, respectively, in statistical signal processing. From 2002 to 2003, he was a Postdoctoral Fellow for six months at Sherbrooke University, Canada. From 2011 to 2012, he was a Visiting Researcher in the SATIE Laboratory, Ecole Normale Suprieure de Cachan, and at the University of Aalborg, Denmark. Since 2003, he has been an Associated Professor in the Laboratory of Signals and Systems (L2S) at Universit Paris-Sud. His teaching activities include basic and advanced notions of statistical signal processing, in particular in the context of the Master2R "Automatique et Traitement du Signal et des Images." He is the author or coauthor of more than 120 publications, has supervised 11 Ph.D students, 4 postdocs, and has been involved in more than 15 collaborative projects including 2 Networks of Excellence projects. His research interests include compressive sampling of nonbandlimited signals, array signal processing, Bayesian performance bounds for parameter estimation and detection, security in mobile networks, as well as numerical linear and multilinear algebra. He received a "Habilitation à Diriger des Recherches" from the Universit Paris-Sud in December 2012. He received the prize for excellence in scientific research "prime d'excellence scientifique" in french. He is an Elected Member of the Commission Consultative de Spécialistes de l'Université in section 61, an Elected Member of the L2S Laboratory Board, and a Nominated Member of the Conseil National des Universités in section 61. He is a Member of the EURASIP Special Area Team Signal Processing for Multisensor Systems and an Affiliate Member of the IEEE Sensor Array and Multichannel Technical Committee. He is also a Scientific Expert for the French National Agency for the research.

Martin Haardt (S'90–M'98–SM'99) studied electrical engineering at the Ruhr University, Bochum, Germany, and at Purdue University, West Lafayette, IN, USA. He received the M.S. degree from Ruhr University, Bochum, in 1991 and the Ph.D. degree from Munich University of Technology, Munich, Germany, in 1996.In 1997, he joint Siemens Mobile Networks in Munich, Germany, where he was responsible for strategic research for third-generation mobile radio systems. From 1998 to 2001, he was the Director for International Projects and University Cooperations in the mobile infrastructure business of Siemens in Munich, where his work focused on mobile communications beyond the third generation. During his time at Siemens, he also taught in the international Master of Science in Communications Engineering program at Munich University of Technology. Since 2011, he has been a Full Professor in the Department of Electrical Engineering and Information Technology and the Head of the Communications Research Laboratory at Ilmenau University of Technology, Germany. Since 2012, he has also been serving as an Honorary Visiting Professor in the Department of Electronics at the University of York, U.K. In the fall of 2006 and the fall of 2007, he was a Visiting Professor at the University of Nice, Sophia-Antipolis, France, and at the University of York, U.K., respectively. His research interests include wireless communications, array signal processing, high-resolution parameter estimation, as well as numerical linear and multilinear algebra. He received the 2009 Best Paper Award from the IEEE Signal Processing Society, the Vodafone (formerly Mannesmann Mobilfunk) Innovations Award for outstanding research in mobile communications, the ITG Best Paper Award from the Association of Electrical Engineering, Electronics, and Information Technology, and the Rohde & Schwarz Outstanding Dissertation Award. He served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING (2002–2006 and 2011–2015), the IEEE SIGNAL PROCESSING LETTERS (2006–2010), the Research Letters in *Signal Processing* (2007–2009), the *Hindawi Journal of Electrical and Computer Engineering* (since 2009), the *EURASIP Signal Processing Journal* (2011–2014), and as a Guest Editor for the *EURASIP Journal on Wireless Communications and Networking*. Since 2011, he has been an Elected Member of the Sensor Array and Multichannel (SAM) Technical Committee of the IEEE Signal Processing Society, where he currently serves as the Vice Chair (since 2015). Moreover, he has served as the Technical Cochair of PIMRC 2005 in Berlin, Germany, ISWCS 2010 in York, UK, as well as the European Wireless 2014 in Barcelona, and as the General Cochair of ISWCS 2013 in Ilmenau, Germany, CAMSAP 2013 in Saint Martin, French Antilles, WSA 2015 in Ilmenau, SAM 2016 in Rio de Janeiro, Brazil, and CAMSAP 2017 in Curacao, Dutch Antilles.